

MAGAZIN FÜR PROFESSIONELLE INFORMATIONSTECHNIK

1 Januar 2008 € 5,50 H 10554

Tutorial:
C++ mit Boost
Thread-Programmierung

E-Mail ohne Spam

Blacklist-Vergleich, Adressen richtig nutzen

Desktop:
KDE 4.0 für Linux und mehr

IT-Sicherheit aus einer Hand:
Unified Threat Management

Server-Technik:
Strom sparen mit Xeon Quad-Core

Spielzeug oder Business-Handy?
Apples iPhone

Betriebssysteme:
Virtualisierung mit Windows Server 2008

Mac OS X 10.5 Leopard

Webentwicklung:
Captchas bauen

Semantic Web:
Query-Sprache SPARQL

Österreich € 6,20 Schweiz CHF 10,70 Benelux € 6,70 Italien € 6,70 iX Magazin für professionelle Informationstechnik www.ix.de Januar 2008

Auf DVD im Heft:
Windows Server 2008
Release Candidate 0

Lauffähig zur kostenlosen Evaluierung bis 7. 4. 2008 – eine Aktivierung ist erst nach 30 Tagen erforderlich.

ANZEIGE

Datenträger enthält Info- und Lehrprogramme gemäß § 14 JuSchG

4 199148 505501 01

Layer-8- und andere Probleme

Dass die Daten der Bürger missbrauchssicher aufgehoben seien, ist eines der Mantras, das die Befürworter von Datensammlungen und Überwachungsausweitungen nicht müde werden zu wiederholen.

Darum war, als vor Kurzem in Großbritannien die persönlichen Daten von 25 Millionen Bürgern verloren gingen, die Empörung auch besonders groß. Vor allem, weil es für den – vermuteten – Datendiebstahl keiner besonderen technischen Finesse bedurfte: Da hatte schlicht jemand zwei CDs ohne den Hauch von Schutzmaßnahmen per Post verschicken wollen.

Solche Fehler werden in Untersuchungsberichten üblicherweise unter „menschlichem Versagen“ subsumiert, ein sogenanntes Layer-8-Problem also.

Deutlich mehr Gehirnschmalz investierten Datendiebe im März dieses Jahres beim Einbruch in die Computersysteme der US-Einzelhandelskette TJX. Dafür gab es auch mehr Daten: 45,7 Millionen Kredit- und Debit-Karten-Nummern von TJX-Kunden.

Dagegen wirkt der Datendiebstahl, der im Oktober beim Hamburger Ticketverkäufer Kartenhaus stattfand, wie kaum noch der Rede wert: Betroffen waren „nur“ die Daten von 66 000 Kunden.

Letztes Beispiel: 2006 benutzten niederländische Hacker kanadische Behörden-PCs mindestens zwei Monate lang zum Verbreiten von Pornos und Raubkopien.

Diese Liste ließe sich beinahe beliebig verlängern, viele Pannen kommen nie an die Öffentlichkeit. Zudem illustrieren solche Fälle, was bei etwas Nachdenken ohnehin klar ist: Daten sind bei Behörden nicht sicherer aufgehoben als bei Privatbetrieben. Menschliches Versagen und technische Unzulänglichkeiten gibt es überall.

Man kann, um auf das Eingangsstatement zurückzukommen, in der Tat unterschiedlicher Ansicht sein über die richtige Balance zwischen persönlicher Freiheit und obrigkeitlichen Einschränkungen, zwischen dem Grad der informationellen Selbstbestimmung und der Reichweite staatlicher Informationssammlung. Wo man sich hier persönlich ansiedelt, hängt von der qualitativen und quantitativen Bedrohungseinschätzung, dem grundsätzlichen Vertrauen in den Staat im Allgemeinen und Besonderen et cetera pp ab.

Eines sollte aber klar sein: Es gibt keinen grundsätzlichen Schutz vor Datendiebstahl. Weil es das Layer-8-Problem immer geben wird. Weil beim Wettrennen zwischen Verschlüsselern und Code-Knackern mal die einen, mal die anderen gewinnen. Weil kein Verschlüsselungsverfahren ewig hält.

Das sollte, wer mehr Daten sammeln möchte, den Betroffenen ehrlich sagen. Und sich lieber über Versicherungen gegen Identitätsdiebstahl und andere Kollateralschäden Gedanken machen, statt seine Zeit mit leeren Versprechungen zu verschwenden.

Jürgen Seeger

JÜRGEN SEEGER



Januar 2008

MARKT + TRENDS

TITEL

REVIEW

REPORT

Mac OS X 10.5 „Leopard“

Lange musste die Apple-Gemeinde auf die neue Version ihres Desktop-Betriebssystems warten. Dafür hat man in Cupertino auch tief in die Feature-Kiste gegriffen und zum Beispiel eine innovative Datensicherungssoftware entwickelt. Was bei Mac OS 10.5.1. neu ist, was schon funktioniert und was noch nicht, ab

Seite 72



Apples iPhone als Business-Handy

Ohne Zweifel exorbitant ist der Coolness-Faktor von Apples iPhone. Die Konfrontation des Livestyle-Gadgets mit den Niederungen des Geschäftsalltags gibt Aufschluss, wie groß die Chancen für eine dauerhafte Begeisterung sind.

Seite 68

Nicht nur für Linux: KDE 4.0

Die grafische Benutzeroberfläche KDE läuft nicht nur auf Linux, sondern auch auf vielen anderen „unixoiden“ Betriebssystemen. KDE 4.0 führt zahlreiche Neuerungen ein, die das freie GUI auf Augenhöhe mit Mac OS X und Windows Vista bringen sollen.

Seite 62





E-Mail ohne Spam

Dass rund 90 % des E-Mail-Aufkommens aus Spam bestehen, ist wegen der konkurrenzlos niedrigen Kosten auf absehbare Zeit nicht zu ändern. Doch sowohl privat als auch geschäftlich kann man durch den geschickten Umgang mit E-Mail-Adressen und die richtige Auswahl von Schutzmaßnahmen dieses Kommunikationsmittel weiterhin effektiv nutzen.

Seite 48 und 56

IT-Sicherheit aus einer Hand

Das Schlagwort „Unified Thread Management" soll signalisieren, dass durch solch eine Lösung gleich mehrere Sicherheitserfordernisse abgedeckt sind. Das heißt aber längst nicht, dass jedes angebotene Produkt den eigenen Bedürfnissen gerecht wird. Die *iX*-Marktübersicht zeigt, worauf man achten muss.

Seite 114



WISSEN

PRAXIS

MEDIEN

RUBRIKEN

Unzulässige Verallgemeinerung

(Editorial: Xid antwortet nicht; *iX* 12/07; S. 3)

Ich bin kein großer Linux-Anwender und kann die fachlichen Aspekte Ihres Editorials nicht beurteilen. Allerdings hat mich Ihr Einleitungssatz „Linux, das stand einmal …“ stutzig gemacht. Besser gesagt: Sie haben mich damit in die Falle gelockt. Denn der Rest des Artikels dreht sich, so weit ich das beurteilen kann, nicht um Linux per se, nur noch um Open Suse 10.3. Mit anderen Linux-Systemen möchten Sie laut eigener Aussage ja gar nicht kämpfen.

Deshalb wäre etwas mehr Sorgfalt angebracht. Denn was Sie schreiben, ist eine unzulässige Verallgemeinerung, die eigentlich nicht stehen bleiben darf.

REINHARD ENGEL,
VIA E-MAIL

(K)Ubuntu existiert!

(Editorial: Xid antwortet nicht; *iX* 12/07; S. 3)

Ihre Erfahrungen mit Linux, die Sie im Editorial der Dezemberausgabe beschrieben haben, kann ich nicht ganz teilen. Hier einige Gedanken dazu:

Der Ratschlag „Nimm doch Ubuntu“ ist hier durchaus nicht ganz fehl am Platz. Meine bisherigen Erfahrungen mit Distributionen wie SuSE und RedHat haben mich dazu gebracht, diese zu meiden wo es geht. Nicht, dass es der Weisheit letzter Schluss wäre, aber meine Gentoo-Installation zu Hause läuft abgesehen von Software-Updates weitgehend stressfrei, ressourcenschonend und die beschriebenen Probleme mit dem NVIDIA-Treiber, den Sound-Daemons, Samba und X.org kann ich nicht nachvollziehen. Außerdem bietet Gentoo ein wunderbares Wiki, ein Forum sowie einen Bugtracker, wo ich bisher noch zu (fast) jedem Problem eine Lösung gefunden habe. Bei Ubuntu ist es ähnlich.

Auch den Vergleich mit Mac OS X finde ich etwas unangebracht. Geschäftlich arbeite ich fast ausschließlich auf Apple-Rechnern und bin eigentlich recht zufrieden mit ihrem Betriebssystem. Trotzdem kämpfe ich ständig mit kleineren bis größeren Reibereien, zu denen die Lösungen (wenn es denn welche gibt) meist schwieriger zu finden sind als für Gentoo.

Linux auf dem Desktop muss nicht problematisch sein. Um es mit Ihren Worten auszudrücken: (K)Ubuntu existiert!

GREGOR RIEPL, VIA E-MAIL

Gemischte Gefühle

(Editorial: Xid antwortet nicht; *iX* 12/07; S. 3)

Mit sehr gemischten Gefühlen habe ich Ihr Editorial in der o. g. Ausgabe gelesen. Leider erinnert es mich an eine vor ein paar Tagen gemachte eigene Erfahrung mit Linux.

Ich arbeite nach Anfängen auf Atari und Mac inzwischen seit langer Zeit unter Windows (95-Vista). Da ich mich sehr für Linux interessiere, habe ich

DER DIREKTE DRAHT ZU iX

Redaktion iX
Postfach 61 04 07
30604 Hannover

Fax: 05 11/53 52-361
E-Mail: <user>@ix.de
Web: www.ix.de

Direktwahl zur Redaktion: 05 11/53 52-387

Für telefonische Anfragen zu Artikeln, technischen Problemen, Produkten et cetera steht die Redaktion wie gewohnt während der Lesersprechstunde zur Verfügung. Und zwar:

Montag bis Freitag, 11 bis 12 Uhr

Bitte nur während der genannten Zeiten anrufen und möglichst die angegebene Durchwahl benutzen.

<Durchwahl>	<user>	
-387	post	Redaktion allgemein
-377	avr	(André von Raison)
-590	ck	(Christian Kirsch)
-387	cle	(Carmen Lehmann)
-374	hb	(Henning Behme)
-379	jd	(Jürgen Diercks)
-386	js	(Jürgen Seeger)
-367	ka	(Kersten Auel)
-153	mm	(Michael Mentzel)
-787	mr	(Michael Riepe)
-373	rh	(Ralph Hülsenbusch)
-689	sun	(Susanne Nolte)
-368	un	(Bert Ungerer)
-535	ur	(Ute Roos)
-384	wm	(Wolfgang Möhle)

Listing-Service:
Sämtliche in iX seit 1990 veröffentlichten Listings sind über den iX-FTP-Server erhältlich:
ftp.heise.de/pub/ix/

unter Virtual Box ein paar Linux-Systeme zum Kennenlernen aufgesetzt. Nachdem ich z. B. bei Ubuntu 7.10 nicht herausfinden konnte, wie ich mich als Admin – Verzeihung, als Root – anmelden konnte, habe ich in einem User-Forum mit einer entsprechenden Frage ganz nett um Hilfe gebeten.

Dass ich nicht buchstäblich gesteinigt wurde, ist auch schon alles. Was ich dort an Antworten lesen musste, war leider an Arroganz nicht mehr zu überbieten. Von „Wenn du das nicht selbst findest, dann lass es lieber." bis „Dir sollte man den Root komplett sperren." war alles dabei. Ein einziger netter Teilnehmer hat mir dann einen Link zukommen lassen, womit ich mein Problem lösen konnte.

Hier liegt leider genau das Problem begraben. Solange Linux (egal welche Distribution) nicht ähnlich einfach wie derzeit Windows XP zu bedienen und dabei gleichzeitig wesentlich stabiler ist, wird es auf Desktops keine größere Verbreitung finden, da die Mehrheit einfach nicht die Zeit und Energie aufbringen wird, sich stundenlang mit Fehlersuche zu beschäftigen. Und unter Windows kennen Sie wenigstens jemand, den man anrufen kann …

BENJAMIN KOHBERG,
VIA E-MAIL

Insgesamt sehr stabil

(Editorial: Xid antwortet nicht; *iX* 12/07; S. 3)

Ihre Kritik an Opensuse 10.3 kann ich (zum Glück) nicht nachvollziehen. Bei mir gab es (fast) keine Probleme beim Umstieg von 10.2 auf 10.3. „Fast" heißt in meinem Fall:
– NFS nutzt jetzt auch Ports im unteren Bereich, so dass ich am Server die entsprechenden Einstellungen machen musste.
– Die ATI-Treiber aus dem Repository sind m. E. nicht zu gebrauchen. Wenn aber die Original-ATI-Treiber geladen und installiert sind, gibt's keine Probleme.
– YP-Bind startet etwas zu spät, so dass ich nach dem Booten den X-Server noch mal starten muss (3-Finger-Griff).

Es sind lästige Fehler, die auch zum Aufspüren viel Zeit benötigen, aber insgesamt läuft das System bei mir sehr stabil.

Inwieweit es Probleme mit aktuellen Systemen und anderen Treibern gibt, kann ich nicht beurteilen. Aus diesem Grund liegt es mir auch fern, Ihre Erfahrungen zu kritisieren. Trotzdem scheint es kein prinzipieller Systemfehler zu sein.

DIETER WINKLER,
VIA E-MAIL

Werbebotschaft

(Editorial: Xid antwortet nicht; *iX* 12/07; S. 3)

Als neuer Abonnent der *iX* hat mich Ihr Editorial sehr verärgert. Erst im letzten Satz vermitteln Sie Ihre Mac-OS-Werbebotschaft. Ich habe bisher die c't als sehr ausgewogenes Magazin geschätzt und nun *iX* für meine dienstlichen Informationen abonniert.

Ich bitte Sie doch zukünftig, redaktionelle Inhalte sauber von Werbung zu trennen. Außerdem erwarte ich von einem „Magazin für professionelle Informationstechnik" Informationen über Software, welche eben genau für die Zielgruppe ausgerichtet ist. Das Editorial sah eher so aus, als hätten Sie mal eben die DVD aus einer der letzten c't-Ausgaben eingelegt und Ihr funktionierendes System damit zerschossen.

Sie erwarten von einer Community-Distribution, dass es auf allen Maschinen für jeden normalen Anwender ohne Probleme einzurichten ist!? Mac OS X ist ein kommerzielles System und wird immer zusammen mit der passenden Hardware ausgeliefert. Es gibt mit Sicherheit Händler, welche ein optimal eingerichtetes professionelles Linux-System komplett mit Hardware zum Preis eines Macs verkaufen möchten. Wie Sie sicher wissen, hat Dell den ersten zaghaften Versuch bereits gestartet.

NORBERT SCHULZE,
VIA E-MAIL

In eigener Sache

Zum Jahreswechsel ändert sich der Dienstleister für die Auslieferung der *iX*-Abos. Bitte entnehmen Sie die neuen Kontaktdaten dem Impressum unter „Abo-Service" (Seite 169). Bis 21. Dezember 2007 gilt noch die bisherige Anschrift: Leserservice *iX*, Postfach 77 71 25, 30821 Garbsen, Tel.: 051 37/ 88 20 00, Fax: 051 37/88 17 12.

Unverändert bleibt die E-Mail-Adresse: abo@heise.de. Bei Anfragen zwischen dem 27.12. und 4.1. kann es zu Verzögerungen kommen, wir bitten, dies zu entschuldigen.

CMS vergessen

(Web Content Management: Fünf CMS auf PHP-Basis; *iX* 12/07, S. 54)

Zuerst einmal Danke, dass Sie Open-Source.CMS testen. Leider vermisse ich in Ihrem Artikel XOOPS (www.xoops.org), welches eine ungemein aktive Community weltweit hat.

WOLFGANG MURTH,
VIA E-MAIL

Wegen der großen Zahl freier Content-Management-Systeme hätte jeder Anspruch auf Vollständigkeit den Rahmen eines Zeitschriftenartikels gesprengt. Wir haben uns darum auf zwei verbreitete Programmiersprachen, Java und PHP, und eine exemplarische Auswahl beschränkt, die die gesamte Spannbreite möglicher Einsatzzwecke heutiger Open-Source-CMS abdeckt. (d. Red.)



Drupal kann mehr

(Web Content Management: Fünf CMS auf PHP-Basis; *iX* 12/07, S. 54)

Sowohl im Artikel als auch in der Vergleichstabelle sind einige Angaben nicht völlig korrekt. Die standardmäßige Template-Sprache von Drupal ist das wohl nicht ganz unbekannte phpTemplate, so dass vorhandene Templates auf dieser Basis ohne große Anstrengung angepasst werden können. Es lassen sich außerdem weitere Template Engines nachrüsten, z. B. phpTal oder Smarty. Auch werden zwei sehr elementare und vielseitig einsetzbare Module nicht erwähnt, die gewissermaßen zum Drupal-Standardrepertoire jeder größeren Seite gehören und die das System von den anderen des Vergleichs absetzen, auch Typo3, nämlich das „Content Construction Kit“ (http://drupal.org/projects/cck) einerseits und „Views“ (http://drupal.org/project/views) andererseits. Weiterhin gibt es, anders als in der Tabelle auf Seite 60 angegeben, sehr wohl eine Sandbox für Drupal (http://drupal.org/project/demo). Anders als bei z. B. Papaya und Typo3 wird auch nicht auf die ausgeklügelten und gut dokumentierten APIs für fast alle Aspekte von Drupal verwiesen.

JAN STÖCKLER,
MÜNSTER

Regeln verletzt

(C++-Programmierung: Boost-Tutorial I; *iX* 12/07; S. 134)

Leider verwendet der Autor den *shared_ptr* genau so, wie man ihn nicht verwenden sollte. In Listing 2: Falls *new payload(„B“)* eine Exception wirft (z. B.: *std::bad_alloc*), wird *objectA* niemals zerstört. Das Beispiel verletzt die grundlegendste Smart-Pointer-Regel: Immer direkt in den Smart Pointer konstruieren, und nur **eine** solche Konstruktion pro Expression (wegen Undefiniertheit der Evaluierungsreihenfolge von Funktionsargumenten). Also folgendermaßen:

```
shared_ptr<payload> A( new payload( "A") );
shared_ptr<payload> B( new payload("B") );
```

Eine zweite Sache: Seit v1.32 kann man *Bind*-Expressions mit Operatoren verbinden, z. B. in Listing 6

```
bind( &payload::getObjectId, _1 ) >
bind( &payload::getObjectId, _2 )
```

statt

```
bind( std::greater<std::string>(),
bind( &payload::getObjectId, _1 ),
bind( &payload::getObjectId, _2 ) )
```

Dies ist ein Feature von *Boost.Bind* (bzw. dessen Superset, *Boost.Lambda*). *std::tr1::bind* erlaubt dies nicht.

MARC MUTZ, BERLIN

Marc Mutz hat Recht. In Teil 2 des Tutorials in diesem Heft findet sich dazu eine nähere Erläuterung. (d. Red.)

Die *iX*-Redaktion behält sich Kürzungen und auszugsweise Wiedergabe der Leserbriefe vor. Die abgedruckten Zuschriften geben ausschließlich die Meinung des Einsenders wieder, nicht die der Redaktion.

Microsofts TechEd: Neues Synchronisations-Framework

Auf der Beta-Bremse

Holger Schwichtenberg

Große Vorstellungen und Ankündigungen waren auf Microsofts europäischer TechEd Mangelware – das neue Sync Framework ragte aus den vielen Kleinigkeiten heraus.

Microsofts europäische Ausgabe der hauseigenen „TechEd“ war hinsichtlich Neuankündigungen in diesem Jahr unspektakulär. Nicht wenige Entwickler begrüßen allerdings die Verschnaufpause im kontinuierlichen Strom neuer Beta-Versionen aus Redmond, der sie in den letzten Monaten in Atem gehalten hatte.

Die zentrale Nachricht in der Keynote von Microsofts Corporate Vice President S. Somasegar war, dass das .Net Framework 3.5 und Visual Studio 2008 schon einige Wochen eher erscheinen als erwartet: „Ende November“. Da der offizielle Veröffentlichungstermin der 27. Februar 2008 ist, lag nahe, dass Microsoft zunächst eine Release-Candidate(RC)-Version verteilen und zum Weihnachtsfest die fertige Version präsentieren würde. Die RC-Version war aber Insidern vorbehalten und seit dem 19. November liegt .Net 3.5 nun in der endgültigen Version auf den Webservern in Redmond.

Am Umfang von .Net 3.5 hat sich nichts mehr geändert. Highlight ist die einheitliche Anfragesprache Language Integrated Query (Linq) für Objekte, Datenbanken und XML-Dokumente. Die Sprachsyntax von C# und Visual Basic haben neben Linq-Befehlen noch einige Neuerungen hinsichtlich funktionaler Programmierung, die Erweiterbarkeit bestehender Klassen und ein paar syntaktische Zuckerstückchen erhalten. Die Windows Communication Foundation (WCF) bietet mehr Protokolle und Formate als bisher sowie eine Integration in die Workflows. Bei Windows Forms, ASP.Net und der Windows Presentation Foundation hat sich wenig verändert. ADO.Net bietet neue Klassen zu Datensynchronisation.

Mehr Features im Service Pack 1

Da sich einige ursprünglich für .Net 3.5 geplante Entwicklungen länger als vorgesehen hingezogen haben, dürften die Redmonder Mitte 2008 schon genug Funktionen für ein .Net 3.7 oder 3.8 zusammenhaben. Aus mehreren Ecken war auf der Veranstaltung jedoch zu hören, dass es wohl ein umfangreiches „Service Pack 1“ für .Net 3.5 werden soll.

Service Pack 1, das aufgrund des Umfangs an Neuerungen eher „Feature Pack“ heißen sollte, enthält neben dem ADO.Net Entity Framework wahrscheinlich die ADO.Net Data Services für den einfachen Zugriff auf Daten über REST-basierte Webservices, neue dynamische Datensteuerelemente für ASP.Net, Unterstützung für die Browsernavigation in ASP.Net Ajax sowie Silverlight-Unterstützung in ASP.Net Ajax.

Ebenfalls auf der Agenda für Service Pack 1 ist MVC, das Model-View-Controller-Prinzip für Webanwendungen. MVC-Seiten sind eine Variante von ASP.Net-Seiten, mit einer Kompetenztrennung in Geschäftsmodell, Präsentation und Ablaufsteuerung. Allerdings plant Microsoft überraschenderweise ein Modell, das nicht auf den bestehenden ASP.Net-Webforms und deren Serversteuerelementen aufsetzt, sondern eins, das an die klassischen Active Server Pages (ASP) erinnert: Im View müssen Webprogrammierer HTML-Elemente explizit verwenden und sie mit Verweisen auf die Daten vermischen. Die Seitenzustandsverwaltung mit dem ASP.Net-Viewstate ist damit in MVC-Seiten ebenfalls nicht verfügbar. Wer eine Tabelle ausgeben will, kann sich dementsprechend nicht auf die Abstraktion des GridView-Steuerelements stützen, sondern muss die HTML-Elemente explizit verwenden und die zugehörige Schleife über die Datensätze programmieren.

Nachdem Webentwickler seit sechs Jahren von dem Rendering der ASP.Net-Serversteuerelemente verwöhnt wurden, ist das ein schlimmer Rückfall in die Websteinzeit. Seiten nach dem MVC-Modell haben in der aktuellen Vorabversion neben der Kompetenztrennung nur einen Vorteil: Sie sind performanter. Insgesamt ist MVC für ASP.Net jedoch eine große Enttäuschung.

Ob die MVC-Vorgehensweise wirklich im nächsten Jahr als Teil des .Net Frameworks erscheint, bleibt deshalb abzuwarten. Auf der TechEd war außerdem zu hören, dass andere schon als Vorabversion herausgegebene Webtechniken aus dem früheren Atlas-Projekt vorläufig gestorben sind. Dazu gehören insbesondere die clientseitige Datenbindung und das sogenannte „XML Script“, das das Konzept der Serversteuerelemente auf den Browser übertragen sollte.

Framework für die Synchronisation

Wirklich neu war das Sync Framework. Dabei handelt es sich um eine Bibliothek zur Synchronisierung von Daten – sowohl zwischen einem Server und einem Client, der zeitweise offline ist, als auch zwischen zwei gleichberechtigten Systemen (Peer-To-Peer). Die Basis (beispielsweise Datenbanken oder Dateisystem) ist durch den Einsatz eines Providermodells beliebig, ebenso das Protokoll

Weitere TechEd-Ankündigungen

Wie im letzten Jahr fand die TechEd Europe über zwei Wochen statt: Die erste richtete sich an Entwickler, die zweite an Administratoren und Infrastrukturexperten. Hier eine Auswahl weiterer Ankündigungen:

– Microsofts Partnerunternehmen (Visual Studio Industry Partner – VSIP) können zukünftig den Quellcode von Visual Studio lizenzieren, um Erweiterungen für die Software einfacher erstellen zu können.

– Für die Mashup-Website Popfly gibt es nun ein Visual Studio-Add-In (Popfly Explorer).

– Wikis und eine Code-Bibliothek, zu denen jedermann beitragen kann, sollen das MSDN-Webportal interaktiver gestalten.

– Microsoft wird den Quellcode einer VS-Variante zur Anpassung des Computerspiels „World of Warcraft“ (in der Sprache Lua) veröffentlichen, um damit die Öffnung von Visual Studio für andere Anwendungsfälle zu demonstrieren.

In der zweiten Woche kündigte Microsoft Folgendes an:

– Windows Server 2008 soll es in acht verschiedenen Versionen geben: Standard, Enterprise, Datacenter und Web sind die Grundversionen. Die ersten drei gibt es in zwei Varianten jeweils mit und ohne die Virtualisierungstechnik Hyper-V (Microsofts Hypervisor). Die achte Version ist die Server-Variante für den Itanium.

– Das Erscheinen von Windows Server 2008 verteilt sich über das gesamte Jahr. Zum offiziellen Feiertermin am 27. Februar soll nur Visual Studio, nicht aber SQL Server und Windows Server tatsächlich verfügbar sein.

– Hyper-V soll als eigenständiger Server verfügbar sein.

– Die 2008er-Version des SQL Server soll sich komplett über PowerShell-Commandlets administrieren lassen, wie dies beim Exchange Server 2007 schon möglich ist.

zur Datenübertragung. Das Sync Framework basiert intern auf klassischem C++-Code, bietet aber außerdem eine .Net-Schnittstelle. Ein Erscheinungstermin steht allerdings noch nicht fest.

Microsoft musste sich viele Fragen anhören, in welcher Beziehung das Sync Framework zu den in .Net 3.5 enthaltenen ADO.Net Synchronisation Services und dem erweiterbaren Synchronisationscenter in Windows Vista stehe. Microsoft versteht das Sync Framework als eine Verallgemeinerung der ADO.Net Synchronisation Services und eine Möglichkeit, Erweiterungen für das Synchronisationcenter in Windows Vista zu schreiben.

Ebenso immer wiederkehrend war die Frage, wieso Microsoft die Lücke im Bereich objektrelationales Mapping (ORM) im .Net Framework nun direkt mit zwei Produkten schließen will: Linq-to-SQL (in .Net 3.5 enthalten) und dem ADO.Net Entity Framework (EF), geplant für das Service Pack 1. Microsoft sieht Linq-to-SQL als Lösung für einfachere Abbildungszenarien und das EF für vollständiges ORM in großen Projekten. Als Begründung für die Konkurrenz im eigenen Lager nannten die Entwickler, dass sich zwei verschiedene Entwicklungsteams (C# und ADO.Net) unabhängig voneinander des Themas ORM angenommen hätten.

PowerShell in der Version 2.0

Jeffrey Snover, Architekt der neuen Windows-Shell, stellte in Barcelona die zweite Vorab-Version der .Net-basierten PowerShell erstmals öffentlich vor. Die wesentliche Änderung betrifft die Fähigkeit zum Fernaufruf von Befehlen und Skripten, die viele Administratoren in der ersten Version vermisst haben. Weiterhin bietet die PowerShell 2.0 Unterstützung für Hintergrundcode-Ausführung, sowie eingeschränkte Shells, die nur bestimmte Befehle und Skripte zulassen. Erstmals will Microsoft einen eigenen Skripteditor anbieten. Derzeit bietet die „Graphical PowerShell“ aber nur Syntaxhervorhebung. Eingabehilfe wie bei Visual Studio ist jedoch geplant.

Darüber hinaus soll es zahlreiche neue Commandlets, neue Operatoren und Verbesserungen in der Unterstützung für die Windows Management Instrumentation (WMI) und die Verzeichnisdienstprogrammierung mit dem Active Directory geben. Bis zur endgültigen Version soll sich noch viel tun. Snover hat noch eine Verpackung für Skripte und zugehörige Datendateien sowie ein Ereignissystem für beliebige Änderungen im System angekündigt. Die PowerShell 2.0 erfordert an einigen Stellen .Net 3.0 und läuft weiterhin auf allen Systemen ab Windows XP.

Die TechEd bot zwar keine weltbewegenden Neuerungen, aber Entwickler konnten die Produktveröffentlichungen der letzten zwölf Monate aufarbeiten – etwa dadurch dass es Vorträge zu ASP.Net 2.0 und eine Einführung in WCF gab, obwohl diese schon lange vorliegen. Erst ab Mitte 2008 gilt wieder: Es ist Beta-Zeit – für .Net 4.0, Visual Studio 10, die nächste Hauptversion von Visual Team System und die neue SOA-Plattform Oslo. Außerdem dürfte 2008 eine Professional Developer Conference (PDC) stattfinden, auf der es traditionell ein Feuerwerk an Neuankündigungen gibt. (hb)

Jülicher Rechner an Spitzenposition unter den Top500

Aufgemotzt

Nikolai Zotow

Trotz aller Kritik am Benchmark Linpack, dem Maß für die Top500, rangeln die Weltbesten nach wie vor um die Plätze. In der November-Liste konnte sich Deutschland mit seinen HPC-Rechenzentren gut behaupten.

Der leistungsfähigste unter den zivil genutzten Supercomputern der Welt steht in Deutschland. In den halbjährlich erscheinenden Top500 der weltweit schnellsten Supercomputer hat der neu installierte Superrechner am Forschungszentrum Jülich den den Sprung auf Platz 2 geschafft. Mit 65 536 Kernen erreicht er eine Spitzenleistung von 167,3 TFlop/s. Es handelt sich um die JUGENE, eine BlueGene/P Solution von IBM mit PowerPC-450-Prozessoren und sagenhaft niedrigen 850 MHz. Nur das militärisch genutzte ältere Modell BlueGene/L am Lawrence Livermore National Laboratory in Livermore, Kalifornien (USA), leistet mit 212 992 Prozessoren mehr: 478,2 TFlop/s. Es stammt einer gemeinschaftlichen Entwicklung von IBM und dem Department of Energy’s (DOE) National Nuclear Security Administration (NNSA).

Überhaupt haben Positionskämpfe unter den ersten zehn die Platzkarten neu gemischt: Auf Platz 3 findet sich als Neuling das New Mexico Computing Applications Center (NMCAC) in Rio Rancho, New Mexico (USA). Es betreibt eine Altix ICE 8200 von SGI in Chippewa Falls, Wisconsin (USA), die mit 14 436 Xeon-Quad-Core-Prozessoren von Intel bei 3 GHZ eine Spitzenleistung von 126,9 TFlop/s auf die Beine stellt.

Überraschungen in den Top Ten

Einsteiger bei den Ländern unter den Top Ten, Indien, folgt auf Platz 4 mit 117,9 TFlop/s. Diese Rechenleistung liefert die Cluster Platform 3000 BL460c von Hewlett-Packard unter Linux. Das in diesem Jahr installierte System ist mit 14 240 Woodcrest-CPUs von Intel bestückt. Die Xeon-Prozessoren sind mit 3 GHz getaktet und über Infiniband gekoppelt. Der Supercomputer steht in Computational Research Laboratories, TATA SONS, in Pune (einst Poonah), Maharashtra (Indien). Das Land konnte bisher in der Juni-Liste 2007 als besten Platz Position 159 belegen und hat inzwischen 9 (vorher 8) Systeme in den Top500.

Als weiterer Neuling taucht der Cluster in Schweden an 5. Position auf. Es handelt sich ebenfalls um die Platform 3000 BL460c von HP, jedoch „nur“ mit 13 728 Xeons bei 2,66 GHz, die 102,8 TFlop/s schafft. Betreiber ist eine Einrichtung der schwedische Regierung (Government Agency). Schweden, im Juni noch mit 10 Systemen und Bestplatz 183, konnte 7 Systeme in der Top500 halten.

Europa hält im kontinentalen Vergleich mit 150 Supercomputern den zweiten Rang mit einem Plus von 23 Systemen. Gemessen an der Rechenleistung konnte Deutschland Großbritannien in der Summe aller Spitzenwerte überholen. Sie hat sich mit 536 TFlop/s gegenüber der letzten Erhebung im Juni 2007 mehr als verdoppelt. Bei der Gesamtanzahl der Implementierungen führt Großbritannien (512 TFlop/s) immer noch mit 48 gegenüber Deutschland mit 31. Es folgt Japan (291 TFlop/s) mit 20 und Frankreich (223 TFlop/s) mit 17. Die USA (4178 TFlop/s) sind nach wie vor Spitzenreiter: 248 der 500 stärksten Rechner stehen dort.

Intel verliert beim Itanium

Spitzenreiter unter den Prozessorherstellern bleibt Intel mit 70,8 %. AMD hält Nummer zwei mit 15,8 % gefolgt von IBMs Power mit 12,2 %. Während Intel beim Xeon seine Position von 231 auf 320 Systeme ausbauen konnte, fiel es mit dem Itanium von 28 auf 21 Installationen zurück. AMD und IBM mussten ebenfalls zurückstecken: 79 (vormals 113) AMD- und 61 (gegenüber 85) Power-Rechner konnte sich in der Bestenliste halten.

Im Vergleich zum letzten Bericht „HP bedrängt IBM – Intel im Plus“ (*iX* 8/07, S. 11) schrumpfte der Vorsprung der einstigen Garagenwerkstatt in Palo Alto. Nur noch 166 Systeme kamen von dort, ein Minus von 44. IBM konnte im gleichen Zeitraum 40 erobern und verbucht jetzt 232 für sich. Dell und SGI folgen weit abgeschlagen mit 24 beziehungsweise 22, konnten aber in den letzten sechs Monaten knapp zulegen. Auf Platz 5 behauptet sich Cray mit 14 (+3). Die restlichen 42 Systeme verteilen sich auf Linux Networx (9), Hitachi (4), Appro International (4), Fujitsu (3) und andere. Bei vier Superrechnern handelt es sich um Eigenbauten.

Jülich ist zweimal vertreten

Die deutsche Top-Liste führt das Forschungszentrum Jülich (FCJ) an, vor dem Leibniz-Rechenzentrum (LRZ) in Garching bei München. Letzteres bietet eine Leistung von 56,5 TFlop/s und nimmt im weltweiten Ranking Platz 15 mit seiner von SGI stammenden Altix 4700 ein, die 9728 Itanium-2-Core mit 1,6 GHz besitzt. Hierzulande folgt auf Platz 3 (Top500 Platz 28) ein weiterer Rechner aus Jülich, der eServer BlueGene mit einer Leistung von 37,3 TFlop/s. Die Max-Planck-Gesellschaft MPI/IPP auf Platz 4 nutzt ebenfalls einen Rechner von IBM, den BlueGene/P. Mit seiner Spitzenleistung von 21,9 TFlop/s kommt er weltweit auf Rang 40. Die folgenden drei Plätze halten nicht näher genannt werden wollende Finanzinstitutionen, weitere Forschungszentren folgen.

Hochleistungsrechnen bleibt eine Linux-Domäne. Das Betriebssystem konnte seine Dominanz gegenüber Juni 2007 ausbauen: auf rund 85,2 % von 77,8 % noch im Juni. Der Anteil an Unix halbierte sich hingegen von 12 % auf 6 %. Einen hohen Zuwachs – wenn auch auf niedrigem Niveau – legte Windows Compute Cluster Server an den Tag. Waren es in der vorigen Statistik magere 0,4 %, kamen nun 1,2 % der RZs, insgesamt sechs Rechensysteme, in die Liste. Der Anteil von gemischten Strukturen fiel von 8,4 % auf 6,8 %. (rh)

Aufs Treppchen: Das Forschungszentrum Jülich konnte mit JUGENE, dem weltweit schnellsten System unter den zivil genutzten Supercomputern, in der Top500 aus dem Stand Platz 2 erreichen.

Handkäs und die Musik

Vom 3. bis 5. Dezember hat es Suns Welttournee, die „Sun Tech Days“, nach Deutschland verschlagen. Gut 1200 Besucher sind laut Veranstalter dem Ruf nach Frankfurt gefolgt.

Zwar gab es einen kostenlosen „Community Day“, allerdings kosteten zwei Drittel der Konferenz Eintritt, regulär 99, ermäßigt 49 Euro. Das ist bei den Sun Tech Days nicht unbedingt üblich und bleibt im Wesentlichen dem Gusto der lokalen Dependance überlassen. Demzufolge fand man in Deutschland, anders als im eintrittsfreien Sankt Petersburg im April dieses Jahres, eher Besucher mit beruflichem Interesse als Studenten.

Die Deutschen erwartete ein recht umfangreiches Programm mit drei Themenschwerpunkten: zwei für Entwickler – Java in sämtlichen Schattierungen und Netbeans sowie Solaris/OpenSolaris. Darüber hinaus gab es Vorträge zu Debugging mit Java, *dtrace* unter Solaris oder zu Performance.

Sun gibt sich „open“

Auf reges Interesse stieß der OpenSolaris-Track. Während der Eröffnung verkündete Frank Curran, dass der Call for Papers für die vom 25. bis 27. Juni 2008 stattfindende OpenSolaris Developer Conference eröffnet sei (www.osdevcon.org). Die GUUG (German Unix User Group) und die CZOSUG (Czech OpenSolaris User Group) organisieren die Konferenz gemeinsam.

Andere Vorträge waren von gemischter Qualität: Scott Rottondo (Sun) benutzte wohl etwas ältere Folien und vergaß beim Vorstellen der Distributionen zunächst Suns „Indiana“. Schade, dass er einige Code-Beiträge externer deutscher Entwickler wie SchilliX, Martux oder ksh93 den Anwesenden nicht auch als hiesige Errungenschaften präsentierte. Das Glanzlicht des Tages setzte Max Bruning – selbstständiger Berater –, der mit viel Gestik zum Besten gab, wie er mit *(k)mdb* bewaffnet einem Bug auf die Schliche kam, bei dem eine Anwendung bei einem *UDP sendto* unreproduzierbar, aber immer zur selben Uhrzeit stehen blieb. Zwei Fehler waren dafür verantwortlich: Die Anwendung (BLOB) übergab einen Sendepuffer von Null, was ein Folgeproblem im Solaris triggerte.

Vater aller Java-Technologien

Zum Auftakt des kostenpflichtigen Konferenzteils dirigierte Games Gosling – der Erfinder von Java und damit der Wegbereiter der Entwicklung des Java-Umfelds – die Aufmerksamkeit von gut 500 Leuten. Die Rede war gut und gewürzt mit Anekdoten: Er berichtete von Handys in Japan, die Java-Spiele mithilfe von 3D-Beschleunigern zu PS3-ähnlicher Performance und entsprechender Begeisterung verhalfen. Der Cola-Automat mit Webserver war eine weitere Anekdote, löste aber bei manchen nur Kopfschütteln aus. Die einzige wirkliche Neuigkeit, mit der Gosling aufwarten konnte, war die Verkündung der Tags zuvor erschienenen Netbeans-Version 6.0, die nun unter anderem mit Unterstützung für (J)Ruby, einer aufgepeppten Oberfläche, einem integrierten Designer für Web- und mobile Anwendungen aufwarten kann.

Im weiteren Verlauf war die Mischung der Vorträge wie zuvor: Uli Gräf von Sun brillierte mit einem Vortrag über ZFS, sein Kollege Scott Rotondos „New Security Features“-Vortrag verweilte zehn Minuten bei *Secure by Default* (nicht mehr Anknipsen aller möglichen Dienste), während er die Trusted Extensions, Labeled Networking überflog.

Unterm Strich: viel Informationen. Bei einigen Vorträgen bestand für technisch stärker interessierte etwas Nachbesserungsbedarf. *Dirk Wetter*



Kölner Mac-Messe ohne Apple und T-Mobile

Zwiegespalten

Achim Born

Herbstgefühle statt Sommer-Jam: Die Terminverschiebung brachte der Maclive Expo keine spürbare Verbesserung. Apple und der hiesige iPhone-Distributor T-Mobile waren erst gar nicht vertreten.

Gänzlich zufrieden konnte nach Ablauf der viertägigen Maclive Expo wohl kaum jemand sein. Denn zum einen mutete das Messekonzept zu widersprüchlich an, zum anderen fanden einfach zu wenige Aussteller und Besucher den Weg in die etwas verborgen liegenden Expo-XXI-Hallen, in denen Mitte November in Köln der Apfeltreff stattfand. Gut 100 Haupt- und Unteraussteller führte der Veranstalter auf.

Die Verschiebung des Termins vom Juni in den November, die bereits im Vorfeld der Veranstaltung einigen Gesprächsstoff geliefert hatte, brachte somit nicht den erhofften Erfolg. Halboffizielle Begründungen – die Verspätung von Mac OS X 10.5 (Leopard) und die iPhone-Ankunft in Deutschland – wirkten wenig überzeugend, denn weder Apple noch T-Mobile waren auf der Maclive Expo vertreten.

Macs kaum zu sehen

Die Abwesenheit von Apple hatte zur Folge, dass neue Rechner, insbesondere Server, nur vereinzelt auf den Ständen von Distributoren – etwa auf dem von Gravis organisierten Macworld-Stand – zu sehen waren. Da es T-Mobile trotz des kurz zuvor erfolgten iPhone-Vertriebsstarts ebenfalls vorzog, der Messe fernzubleiben, konnten die Besucher ihr haptisches Interesse nur an den wenigen Exemplaren am Stand „Meet the iPhone“ von M&M stillen.

Mit dem neuen MacBusinessPark, den Expomedia in Zusammenarbeit mit dem Internet-Portal OSXpert.biz für Besucher aus Unternehmen organisiert hatte, mühte man sich, der Messe insgesamt einen professionelleren Anstrich zugeben. Server- oder Netzwerk-Lösungen waren jedoch nur bedingt bei Distributoren zu begutachten. Brainworks etwa hatte neben der Client-Managementsoftware Lanrev den Mailware- und Groupware-Server Kerio dabei, der mittels Repellent-Funktion durch leichte Verzögerungen des SMTP-Ablaufs ungeduldige Spam-Zombies ausgrenzen soll.

Agenturmitarbeiter oder professionelle Kreative konnten häufiger fündig werden. So gaben wie im Vorjahr auf der Maclive-Bühne bekannte Apple-Kenner Einblicke in ihr Sujet. Die kostenpflichtigen Seminare und Tutorials hatten ebenfalls ihren Zulauf. HP, mit einem 125 m^2 großen Stand vertreten, führte großformatiges Drucken für Profifotografen und Grafikdesigner vor. Microsoft warb außer für das kommende Office 2008 für Mac insbesondere für die Digital-Asset-Management- und Katalogsoftware Expression Media. Einen weiteren Schwerpunkt bildete die Adobe-Flash-Konkurrenz Silverlight, ein Plug-in für Rich Interactive Applications (RIAs) und .Net-basierte Multimedia-Anwendungen, das in allen gängigen Browsern unter Windows und Mac OS X läuft.

Insgesamt wirkte die Messe – auch aufgrund des Fehlens bedeutender Hersteller – etwas beliebig. Sie war weder eine Fach- noch eine Publikumsmesse. Die erstmals vorgenommene Zweiteilung des Messeverlaufs mit je zwei Tagen für Fachbesucher (schwach besucht) und für Mac-Endverbraucher (besser frequentiert) unterstrich den widersprüchlichen Charakter der Veranstaltung. (un)

Löst das IGF die großen Fragen der Netzpolitik?

Eine-Welt-Laden

Monika Ermert

Noch herrscht Uneinigkeit darüber, wie viel Netzpolitik das Internet Governance Forum (IGF) überhaupt beackern soll. Die Themenliste der unter dem Dach der Vereinten Nationen gegründeten Organisation ist beim zweiten Treffen in Rio de Janeiro jedenfalls gewachsen.

Zum Datenschutz, zur Nachhaltigkeit freier Standards für die Entwicklungsländer, zur Balance zwischen Sicherheit und Freiheit kam diesmal das Thema Schutz von Kindern und Jugendlichen im Netz – neben vielen anderen. Das wirft Fragen auf, ob das ursprüngliche Ziel, die Entmachtung der USA bei der Aufsicht über die zentralen Rootserver und das DNS, in diesem Forum überhaupt erreichbar ist.

Der Erfolg einer Mammutveranstaltung wie dem IGF mit rund 1500 Teilnehmern, über 80 Workshops und Hunderttausenden von Seiten Protokollen ist schwer zu messen. Entscheidungsbefugnisse haben Regierungen, die das IGF ins Leben riefen, der neuen Institution nicht verliehen – und dabei müsse es bleiben, unterstrich der Vertreter des US State Department, David Gross, in Rio. Diplomatische Verhandlungen um die Aufsichtsmacht bei den zentralen Rootservern etwa will die USA um jeden Preis verhindern. Er sehe „derzeit keinen Änderungsbedarf“, sagte Gross.

Schlüssel in den USA

Als politisch heikel bezeichnete Denic-Chefin Sabine Dolderer Überlegungen zum raschen Signieren der Rootzone per DNSSec durch die von den USA beaufsichtigte Internet Corporation for Assigned Names and Numbers (ICANN). Mit ihr würde auch das zentrale Schlüsselmanagement ausgerechnet wieder unter die Aufsicht der so heftig kritisierten USA fallen.

Trotzdem seien konkrete Erfolge zu verzeichnen, fand Matthew Sheers von der Internet Society, und verwies auf ein konkretes Beispiel. Mehrere Regierungsvertreter hätten nämlich Interesse gezeigt am Installieren von Anycast-Rootserver-Instanzen und eigenen Internet Exchanges in ihren Ländern. Darüber kann zunächst einmal lokaler Datenverkehr auch auf lokalen Servern gehalten werden. An teurer internationaler Konnektivität, die die Entwicklungsländer in der Regel von großen US-Backbone-Providern einkaufen, lässt sich damit sparen.

Viel zu wenig ist gerade beim Aufbau lokaler Exchanges nach der ersten IGF im vergangenen Jahr geschehen, so Bill Woodcock, CTO beim Packet Clearing House. Das Bewusstsein um die Notwendigkeit eigener Infrastruktur wachse aber etwa bei afrikanischen Regierungen mit jeder IGF, freute sich Nii Quaynor, afrikanischer Internet-Pionier und Leiter der Registry für Ghana. Der gleichberechtigte Dialog zwischen Regierungen, Nichtregierungsorganisationen, Techies und Unternehmen gilt vielen als gelungenes Experiment.

Bei Spezialfragen wie dem Datenschutz könnte sich das IGF als Impulsgeber erweisen. Simon Davis vom der Organisation Privacy International brach in Rio eine Lanze für das ungeliebte Digital Rights Management (DRM) als mögliche technische Lösung zur Sicherung der Privatsphäre im Web 2.0.

IGF BRAZIL 2007 INTERNET GOVERNANCE FORUM

Jan Schallaböck vom Unabhängigen Landeszentrum für Datenschutz Schleswig-Holstein begrüßte den Vorstoß. Per DRM lasse sich das Grundprinzip des Datenschutzes, die Zweckbindung, sehr gut abbilden. Zudem erhalte der Nutzer die Kontrolle darüber, was er tatsächlich dem Gegenüber erlaube. Ein Check der Prüfsumme auf dem dedizierten Server des Datenverarbeiters soll automatisiert sicherstellen, dass keine anderen Operationen stattfinden. Das EU-Datenschutzprojekt PRIME diskutiere bereits solche Lösungen, die Datenschützer leicht überprüfen könnten, so Schallaböck.

DRM für den Datenschutz

Man solle den DRM-Systemen eine Chance geben, stieß Privacy-Experte Davis ins gleiche Horn: „Viele Verfahren sind anfangs komplex, aber mit der Zeit und durch weitere Verbreitung werden sie anwenderfreundlicher und irgendwann sogar unsichtbar. Verschlüsselungslösungen sind ein gutes Beispiel dafür.“

Die IGF-Koalition zu Privacy ist die größte von zehn solcher „dynamischen Koalitionen“, über 70 Unternehmen, Datenschutzbehörden, Universitäten und Nichtregierungsorganisationen sind mit von der Partie. Die gestarteten Projekte umfassen neben den Überlegungen zum Privacy-DRM auch Arbeiten zu Datenschutz und Überwachung und eine Art Creative Commons System für Datenschutzstandards.

Bei der kleineren dynamischen Koalition „Offene Standards“ verwies man in Rio besonders auf die Fortschritte, die man seit dem ersten IGF in Athen auf internationaler Ebene gemacht habe. Offene Standards wurden in die im September von einigen Regierungen verabschiedete entwicklungspolitische Agenda der World Intellectual Property Organisation (WIPO) aufgenommen. Eine Harmonisierung des Patentrechts weltweit hat die WIPO dagegen erst einmal abgelehnt, berichtete Thiru Balasubramaniam von Knowledge Ecology International (KEI). Stattdessen solle der zuständige WIPO-Ausschuss Standing Committee on Patents (SCP) bis zum kommenden Frühjahr einen Bericht vorlegen, der die Kosten des Patentsystems für Entwicklungsländer beziffert.

Nach Rio hatten die Free-Software-Vertreter den Anwalt des European Committee on Interoperable Standards (ECIS), Thomas Vinje, geladen. Er berichtete von der zweiten Untersuchung der EU-Wettbewerbskommission gegen mögliche wettbewerbsfeindliche Praktiken des Softwareriesen Microsoft. Darin gehe es vor allem um die Privilegierung von Outlook, Office und .Net. Der aktuelle Streit um die ISO-Standardisierung von Microsofts OOXML habe einen gewissen Einflss auf das Brüsseler Verfahren, meinte Vinje.

Offene Standards überprüfbar

Ein praktisches Projekt stellte Georg Greve von der Free Software Foundation Europe (FSFE) vor. CertifiedOpen soll künftig öffentlichen Verwaltungen erlauben, vor dem Einkauf von Software und Diensten zu prüfen, ob die ins Auge gefassten Angebote auf offenen Standards basieren. Über CertifiedOpen können sich kommerzielle Anbieter für 400 Euro pro Produkt zertifizieren, nichtkommerzielle Anbieter nutzen den Dienst kostenlos. Die vorgesehene Selbsteinstufung lasse sich durch die Veröffentlichung der von den Anbietern ausgefüllten CertifiedOpen-Fragebögen kontrollieren.

Die FSFE hat mit Certified-Open eine erste Antwort auf die Frage gegeben, die die dynamische Koalition „offene Standards“ beim IGF noch beantworten muss: Welche Kriterien müssen erfüllt sein, damit man überhaupt von einem offenen Standard reden kann? Man möchte gerne eine von möglichst vielen Parteien getragene Definition, daran werde noch gefeilt. Im kommenden Jahr, beim nächsten Treffen in Delhi, dürfte es sie geben. (un)

11 Jahre „Moderner Staat“

Bundeskunden

Barbara Lange

Die Verwaltungen müssen sich umstellen, denn die Zukunft heißt „One-Stop-Government“: Bald wird es nur noch einen Ansprechpartner für EU-Dienstleistungen und eine Behördenrufnummer 115 für alles geben.

Europaweit soll der in Fachkreisen sprichwörtliche portugiesische Friseur ab Ende 2009 seine Dienstleistung mithilfe eines einzigen Ansprechpartners anmelden können. Außerdem müssen die Behörden das gesamte Verfahren elektronisch abwickeln – eine Aufgabe, die den Behörden derzeit auf den Nägeln brennt, und der man sich mit großer Demut nähern sollte, so Harald Lemke, Staatssekretär im Hessischen Ministerium der Finanzen, auf einer Podiumsdiskussion im Rahmen der 11. Fachmesse und des Kongresses „Moderner Staat“ Ende November auf dem Berliner Messegelände.

Auch die Einführung der bundesweit einheitlichen Behördenrufnummer 115, die ab 2008 in einigen Modellregionen getestet werden soll, heißt One-Stop-Government. Hier sollen Bürger – man will sie demnächst Kunden nennen –, Unternehmen und Institutionen künftig alle Ämter des Landes unter einer zentralen Einwahlnummer erreichen können. Während es für Unternehmen und Bürger eine große Erleichterung ist, wenn sie für ein Anliegen nicht mehr an mehrere Amtsstuben klopfen müssen, widerspricht diese Zentralisierung der föderalen Verwaltungsstruktur, in der Behörden aus Bund, Ländern und Kommunen überwiegend isoliert vor sich hin werkeln.

Daher gilt die Umsetzung beider Vorgaben – der EU-Dienstleistungsrichtlinie (EU-DLR) und der Behördenrufnummer 115 – als große Herausforderung, und für die Zusammenarbeit der einzelnen Behörden entsteht ein immenser Regelungsbedarf, ein Umbau der Verwaltung, für den die Verantwortlichen derzeit Pläne mit Zuständigkeiten et cetera erarbeiten.

IT-Blaupause bis 2008

Die Redner der Podiumsdiskussion zeigten sich skeptisch, ob diese Anforderung im geforderten Zeitrahmen umzusetzen ist, angesichts des Gefühls, mit Latschen vor einer Bergwand zu stehen (Lemke).

Eine große Rolle bei der Vereinheitlichung dieser heterogenen Prozesse spielt die IT. Ein technisches Konzept für die IT-Infrastruktur und die Rahmen-Architektur der EU-DLR will man als „Blaupause“ bis Ende 2008 entwickelt haben. Ziel ist eine medienbruchfreie Verfahrensabwicklung mit hersteller- und produktneutralen Lösungen. Dabei kann man aber nicht davon ausgehen, dass im Prinzip alle IT-Informationssysteme schon da sind, die man einfach nur neu zusammenfügen muss, betonte Marianne Wulff, Kommunale Gemeinschaftsstelle für Verwaltungsmanagement (KGSt).

Insgesamt aber war die Stimmung der Branche positiv: Oft hörte man den Verweis auf den E-Government-Statusbericht der EU-Kommission, wonach Deutschland auf Platz 10 vorgerückt sei. Der Internet-Auftritt des Bundestages (www.bundestag.de) wurde sogar kürzlich von der UNESCO als beste E-Government-Website der Welt ausgezeichnet.

Und es gab mit über 3500 (300 mehr als im Vorjahr) Teilnehmern einen Besucherrekord, wie Veranstalter Reed Exhibitions meldete. Über 70 Prozent der Teilnehmer waren Entscheidungsträger aus Bund, Ländern und Kommunen.

Darüber hinaus registrierten die 200 Aussteller von A wie Accenture über M wie Microsoft bis Z wie ZIVIT ein verstärktes Interesse der Besucher an E-Government, Personalmanagement und dem Dauerbrenner IT-Sicherheit, der wie schon im letzten Jahr als Themenschwerpunkt mit Firmenvorträgen und einer vom BSI gestalteten Kongress-Session gut abgedeckt war.

Dort gaben Referenten aus dem BSI einen Überblick über ihr Amt, die Sicherheitslage, die „Sichere Inter-Netzwerk-Architektur“ (SINA) und über die virtuelle Poststelle des Bundes, die die elektronische Kommunikation zwischen Behörden und externen Partnern kryptografisch absichern soll. Derzeit arbeitet man hier an einer Out-of-the-Box-Lösung, die auch kleinere Behörden mit geringen IT-Kapazitäten vorkonfiguriert einsetzen können.

Auto online anmelden und CO_2 reduzieren

Vorgestellt wurde auch das Deutschland-Online-Projekt zum Kfz-Wesen. Bürger alias Kunden und Unternehmen sollen in Zukunft Fahrzeuge online vom PC aus anmelden können – um Kosten zu sparen, aber auch, damit es diese Infrastruktur aufgrund der demografischen Entwicklung auch in Zukunft noch flächendeckend gibt, so Detlef Gottschalck, Staatsrat der Hansestadt Hamburg. Darüber hinaus verfahren potenzielle Autohalter schätzungsweise eine Milliarde Kilometer im Jahr allein für die Zulassung.

Einen großen Stellenwert bekamen auch Themen jenseits der reinen Technik, hier besonders die Podiumsdiskussion „Neue Wege in der Personalführung – Braucht die Verwaltung eine neue Führungskultur?“, an der unter anderem Schirmherr Bundesinnenminister Schäuble gleich nach seiner Auftaktrede teilnahm.

Der 12. „Moderne Staat“ findet am 4. und 5. November 2008 in Berlin statt. Angekündigt sind größere Messehallen und eine noch stärker themenorientierte Struktur. (ur)

DeepSec 2007: Erfolgreiches Debüt in Wien

„Bringing together the world’s leading security professionals from academics, government, industry and the underground hacking community.“ Unter diesem Motto öffnete die Sicherheitskonferenz DeepSec (deepsec.net) im November in Wien zum ersten Mal ihre Pforten. Mariano Nuñez Di Croce referierte in seinem Tutorial über die (Un-)Sicherheit von SAP-Systemen. Er zeigte einen sogenannten Evil-Twin-Angriff auf SAP-Server, bei dem gezielt Remote Function Calls (RFCs) über das System des Angreifers umgeleitet wurden.

In seiner Keynote am ersten Tag erläuterte Paul Simmonds vom „Jericho Forum“, wie überflüssig die klassischen Schutzmaßnahmen an den Netzwerkgrenzen seien. In Zeiten, in denen immer mehr Löcher an dieser Grenze existieren und bei vielen Unternehmen nahezu mehr externe als interne Mitarbeiter Zugang zum Netzwerk haben, stellt sich immer öfter die Frage nach der Sinnhaftigkeit eines solchen Schutzes. Vielmehr müsse sich der Schutz nicht auf die Geräte oder Netze, sondern auf die dort verarbeiteten Daten konzentrieren.

Lukas Grunwald referierte über Designschwächen beim elektronischen Reisepass (ePass). Einen weiteren Beitrag zum Thema RFID lieferte Melanie Rieback, die mit „RFID Guardian“ eine Art aktiven Schutzschild für die Funktechnologie vorstellte. RFID Guardian soll – mitgeführt als Gerät in der Größe eines Mobiltelefons – das unbefugte Auslesen von RFID-Tags des Trägers verhindern. Weiterhin soll es Leseversuche protokollieren und somit auch als eine Art Intrusion-Detection-System (IDS) für RFID einsetzbar sein. Angesichts der erfolgreichen Veranstaltung darf man erwarten, dass auch im kommenden Jahr wieder eine aufschlussreiche DeepSec stattfindet.

Marco Lorenz

Interop für europäischen Markt:

Rundumschlag

Barbara Lange

Mit einem großen Themenspektrum von IT-Sicherheit bis zur Breitbandversorgung trat die Konferenz und Ausstellung Interop zum ersten Mal auf dem Berliner Messegelände an. Vielleicht auch zum letzten Mal?

Auf der Fachtagung IT-Sicherheit gaben Vertreter von McAfee, Eleven und Retarus einen Überblick über die Bedrohungslage. Demnach sind nach wie vor E-Mails der größte wunde Punkt der IT-Sicherheit, und Spam ist nach Ansicht aller zu einem knallharten lukrativen Geschäftsmodell geworden. Zum Beispiel Aktien-Spam. Als Folge derart „beworbener“ Wertpapiere steigen die Kurse kurzfristig stark an, zeigte Andreas Beierer von Eleven.

Mittlerweile sind nur noch klägliche fünf bis sieben Prozent des E-Mail-Aufkommens als geschäftsrelevant zu bezeichnen, während der über 90-prozentige, zunehmend über Botnetze verbreitete Müll die Stabilität der IT-Infrastrukturen gefährdet. Dabei haben seriös aussehende PDF-, Excel und ZIP-Anhänge das Spam-Spektrum erweitert.

BSI: Gezielte Angriffe nehmen zu

Einen Anstieg der Wirtschaftsspionage über IT-Wege beklagte Udo Helmbrecht, Präsident des Bundesamts für Sicherheit in der Informationstechnologie (BSI) in seinem Vortrag „Status Quo der IT-Sicherheit in Deutschland“. Besondere Sorge bereiten ihm gezielte Angriffe, die Unternehmen und Behörden lahmlegen. Das Sicherheitsbewusstsein in den öffentlichen Verwaltungen sei nicht ausreichend, kritisierte er. Derzeit arbeiten BSI und Wirtschaft an der Umsetzung des „Nationalen Plans zum Schutz der Informationsinfrastrukturen“.

Gut besucht war das Live-Hacking von Sebastian Schreiber von Syss GmbH, der zeigte, wie sich systemimmanente Schwachstellen auswirken können: Da wurde ein trojanerverseuchtes Handy zur Abhörwanze und gab den Aufenthalt des Handy-Besitzers bekannt, ein Babyphone spähte drahtlose Überwachungskameras aus. Mit einfachen URL-Manipulationen im Warenkorb von Onlineangeboten korrigierte er den Preis nach unten. E-Commerce ist und bleibt unsicher, so sein Fazit.

Interop – wie weiter?

Mit der weltweit in Las Vegas, New York, Tokio, Moskau und Sao Paulo und Anfang November zum ersten Mal in Berlin ausgerichteten Interop wollten die Veranstalter die spezifischen Bedürfnisse des europäischen Marktes ansprechen. Mit 10 000 Besuchern hatte man gerechnet. So viele waren es sicher nicht – wie viele genau, bleibt das Geheimnis der Veranstalter, die bis Redaktionsschluss drei Wochen nach dem Event trotz mehrmaligen Nachfragens keine Auskünfte über Teilnehmerzahl und Fortsetzung geben konnten.

Vielleicht war das Themenspektrum so kurz nach der Münchener Systems zu groß. So gab es neben der Fachtagung IT-Security einen Fachkongress zur Planung und Implementierung von IP-Telefonie-Infrastrukturen, eine Fachtagung Wireless & Mobility, die sich mit der Verknüpfung von Mobilfunk und WLAN beschäftigte, den „Baltic Broadband Day“ über die Versorgung von ländlichen Regionen in Europa mit Breitbandanschlüssen und die Microsoft-Fachtagung „Application Interoperability“. Dazu Keynotes von Avaya, Cisco, Citrix, Hewlett-Packard und Microsoft sowie eine Ausstellung mit 140 Unternehmen. (ur)

PHP-/Ajax-Konferenz: Von Agilität bis Mashups

Familiär

Stefan Priebsch

Dass das Potenzial der Skriptsprache PHP noch nicht erschöpft sei, war eine der guten Nachrichten auf der PHP-Konferenz im November.

Zum zwölften Mal fand Anfang November in Frankfurt-Mörfelden die International PHP Conference (IPC) statt – gleichzeitig die dritte „Ajax in Action“, die die Open Source Database Conference als ständigen Begleiter der IPC abgelöst hat (zusammen 350 Besucher aus 22 Ländern). Zwei Tage vor dem Beginn der Hauptkonferenz liefen Workshops; auf das größte Interesse stießen agile Entwicklungsmethoden sowie das Testen mit PHPUnit und Selenium. Weitere Themenschwerpunkte waren die Qualitätssicherung, das Testen sowie Werkzeuge und Automation.

Ausblicke auf die Zukunft gaben die beiden Typo3-Entwickler Robert Lemke und Karsten Dambekalns. Sie zeigten mit einem Vortrag zur aspektorientierten Programmierung und einem Erfahrungsbericht zur Implementierung eines JSR-283 Content Repository in PHP, dass das Potenzial der Sprache noch nicht erschöpft ist.

Hin zu Aspekten und Content Repository

Ajax in Action deckte eine große Themenbandbreite von den Grundlagen der Entwicklung über das Erstellen einer Offline-Anwendung bis hin zu Mashups und Usability-Tests ab. Dabei kamen verschiedene Techniken wie Ruby on Rails, Silverlight, Flex und ASP.Net zu ihrem Recht.

Nicht nur die Sessions, sondern auch die Keynotes waren mit prominenten Sprechern besetzt. Der Kanadier Zak Greant stellte unter dem Titel „The Age of Literate Machines“ die These auf, dass durch den Einsatz von nicht quelloffener Software im Rechtssystem die Transparenz von Gerichtsurteilen verloren geht, und dies einen Rückschritt gegenüber der Dokumentation auf Papier bedeutet. Für die Zuhörer drängte sich eine gedankliche Parallele zu Wahlmaschinen und -stiften förmlich auf.

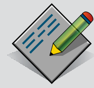
Ein wirkliches Manko der Konferenz sind die immerwährenden Schwierigkeiten mit der schlechten Internetanbindung. Deshalb soll sich 2005 sogar die Release einer PHP-Version verschoben haben, da das Gros der Entwickler auf der Konferenz zu Gast war und nur eingeschränkten Onlinezugang hatte. Dafür gibt es auf kaum einer Konferenz wie dieser einen so offenen Austausch zwischen den PHP-Entwicklern und der Community. Kein Wunder, dass die Konferenz mittlerweile als „Familientreffen“ gilt. (hb)

CMS: Days Communiqué und Alfresco mit Web-2.0-Anbindung

Day Software (www.day.com) stellt mit ihrem CMS Communiqué Avanced Collaboration (CQ AC) eine Variante zur Verfügung, die die Verwaltung sogenannter Social Media erlaubt: Wiki, Blog und Kalender – mit Ajax-Mitteln realisiert. Für Wiki und Blog existiert ein Rich-Text-Editor, Inhalte indiziert Communiqué fortlaufend (Volltextsuche), und Anwender können sich über Wiki-Änderungen per E-Mail informieren lassen.

Mit neuen Werkzeugen für ihr gleichnamiges Enterprise Content Management System (ECMS) will Alfresco Software Web-2.0-Dienste und Tools anbieten (www.alfresco.com). Zu den Integrationskandidaten gehören Facebook, iGoogle, Adobes Flex und Wordpress. Die sogenannte Social-Computing-Plattform enthält laut Alfresco ab Mitte Dezember Tools für die Anbindung der genannten Web-2.0-Vertreter. Das ECMS selbst soll außer der Einbeziehung von Facebook et al. ermöglichen, Inhalte für externe Anwendungen wie Media Wiki oder Wordpress zu erzeugen und mit Flex Mashups zu entwickeln.

KURZ NOTIERT



Widgets: Mit Version 4.5 seiner Widget Engine unterstützt Yahoo in seinen Miniprogrammen Flash und HTML. Außerdem verzichtet Yahoo auf die Bezeichnung „Engine“ und nennt seine Werkzeugkiste nur noch Yahoo Widgets (widgets.yahoo.com/upgrade/).

Wissensmanagement: Meta-Level aus Saarbrücken (www.meta-level.de) hat in der Version 4.4.0 ihres webbasierten Dokumenten- und Wissensmanagementsystems Meta-Dok die Suchfunktionen erweitert und den Bereich des Administrators überarbeitet.

CMS: E-Spirit (www.e-spirit.de) hat sein CMS FirstSpirit in der Version 4 vorgestellt, deren Neuerungen vor allem in umfangreichen Projekten zum Tragen kommen sollen. Unter anderem hat der Hersteller die Versionierung überarbeitet. Außerdem hat er für Redakteure neue Funktionen eingebaut.

Wikipedia vs. Brockhaus

Ob Wikipedia, die von vielen Freiwilligen erstellte Online-Enzyklopädie, gegen die ständig aktualisierte Onlineversion des Brockhaus bestehen könne, wollte der „Stern“ wissen und ließ anhand von 50 zufällig ausgewählten Stichwörtern beide Nachschlagewerke vom Wissenschaftlichen Informationsdienst (Köln, www.wind-gmbh.com) untersuchen. Die Einträge waren thematisch breit gestreut – von Politik über Sport und Wirtschaft bis Unterhaltung.

Beim Wettstreit zwischen Community und Profis schnitt Erstere erheblich besser ab. Mit einer Gesamtnote von 1,7 lag Wikipedia deutlich vor dem Brockhaus mit 2,7. Bei 43 der 50 Artikel bekam Wikipedia die bessere Note, sechsmal lag der Brockhaus vorn; einmal gab es ein Unentschieden. Vor allem hinsichtlich der Aktualität überzeugte die Community: Beim Todesdatum von Luciano Pavarotti oder der Nobelpreisverleihung an Doris Lessing war Wikipedias Eintrag im Gegensatz zu dem des Brockhaus jeweils am selben Tag aktualisiert. Zwar lag die Profi-Enzyklopädie vorn, was Verständlichkeit angeht, aber selbst bei Korrektheit hatte Wikipedia die Nase vorn.

Vor zwei Jahren hatte die US-Zeitschrift „Nature“ die englischsprachige Wikipedia mit der Encyclopedia Britannica verglichen und war nach einem Vergleich von 42 Artikeln aus beiden zu dem Schluss gekommen, dass die Encyclopedia Britannica zwar besser als Wikipedia sei – aber nicht viel.

Googles Handyplattform Android

Wie angekündigt (*iX* berichtete), hat Google im November eine erste Version seines Software Development Kit für die geplante Handy-Software-Plattform Android zum Download freigegeben. Sie basiert auf dem Linux-Kernel 2.6 und enthält außer einer Java-VM weitere Software: von der *libc*-Bibliothek über die aus dem KDE-Projekt stammende Browser Engine WebKit und eine 2D-Grafik-Engine bis hin zu SQLite.

Google forciert die Entwicklung von Android-Anwendungen und hat im November mit einer Developer Challenge genannten Aktion 10 Mio. US-Dollar für Anwendungen ausgelobt, bei der es zwischen 25 000 und 275 000 $ zu gewinnen gibt (siehe code.google.com/android/adc.html).

iPhone als Visualisierungs-Endgerät

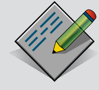
Topix, Anbieter der gleichnamigen, für Mac und Windows verfügbaren Unternehmenssoftware, hat auf der Maclive Expo (siehe S. 17) erstmals die Komponente Analytik-Report vorgestellt. Damit lassen sich Auswertungen von Auftrags- und ähnlichen Daten bereichs- und zeitübergreifend erstellen und visualisieren. Die Ottobrunner Firma präsentierte zudem, wie sich das iPhone als nützliches Endgerät in die Unternehmenssoftware einbinden lässt. In der Grundversion des iPhone-Clients Topix:5 (im Beta-Status verfügbar) lassen sich vertriebsrelevante Adressen, Termine, Aufgaben, Wiedervorlagen et cetera bearbeiten. Der Umfang geht deutlich über das gewöhnliche Adress- und Terminmanagement des iPhone hinaus. In Vorbereitung sind zudem in weiteren Ausbaustufen des iPhone-Clients Funktionen zum Jobmanagement einschließlich Arbeitszeiten und Reisekosten. Selbst die direkte Angebotserstellung und ein Überblick über die gesamte aktuelle Auftragsabwicklung sollen künftig über das neue Apple-Gerät möglich sein.

Topix basiert wie weitere in Köln vorgeführte Produkte (etwa die Warenwirtschaft Business Open von Gubus, die CRM-Lösung deLuxe von Fuchs EDV und Masterfinanz) auf 4D. Der Hersteller selbst gewährte erste Einblicke in die nun verfügbare Entwickler-Produktreihe der Version 11. Kern ist eine Datenbank-Engine, mit der sich die zuvor eher geschlossene Plattform erstmalig für eine direkte SQL-Unterstützung öffnet. Traditionelle, in der 4GL von 4D geschriebenen Anwendungen sind in dieser Version noch lauffähig. Mit der Version 12, die den vollständigen Wechsel zur Objektorientierung dokumentieren soll, scheint dies nicht mehr der Fall sein. *Achim Born*

Entwurf für neue Musterwiderrufsbelehrung

Nachdem mehrere Gerichtsurteile die gesetzliche Musterwiderrufsbelehrung aus der BGB-Informationspflichtenverordnung für teilweise rechtswidrig erklärt haben, hat das Bundesjustizministerium im November einen Entwurf zu einer neuen Fassung vorgelegt. Strittig in der alten Version waren vor allem Beginn und Dauer der Widerrufsfrist. Zwar sind in der Neufassung Teile der kritisierten Passagen umgeschrieben, doch löst sie längst nicht alle Probleme. Insbesondere monieren Juristen die ausgeweiteten Informationspflichten, denen zufolge ein Onlinehändler seinen Kunden zahlreiche Vorschriften aus dem BGB und der BGB-InfoV im Volltext mitteilen muss. Eine rechtsgültige Widerrufsbelehrung müsse nach diesem Entwurf nicht weniger als 12 000 Zeichen umfassen, kritisiert Joerg Heidrich, Justiziar des Heise-Zeitschriftenverlages. Das entspricht fast zwei Druckseiten in einer Heise-Publikation. Der Entwurf ist unter www.bmj.bund.de/files/-/2550/ einzusehen.

KURZ NOTIERT



Geschäftsmodell: Ein Team der Georg-August-Universität Göttingen hat den erstmalig durchgeführten Bitkom-Hochschulwettbewerb „Beste Prozessarchitektur“ gewonnen. Auf Platz zwei folgt ein gemeinsames Team der Uni Potsdam und des Hasso-Plattner-Instituts, Platz drei erklommen Studenten der Leibniz-Uni Hannover.

Auftrieb: Für 2008 prognostiziert das Marktforschungsunternehmen Gartner in Deutschland einen Zuwachs von 5,9 % bei den Ausgaben für Unternehmenssoftware. Die Analysten rechnen mit Umsätzen in Höhe von 2,2 Mrd. Euro für CRM-, ERP- und SCM-Anwendungen.

Model Driven Architecture: MID modelliert das Wachstum

Modellierung soll den Software-Entwicklungsprozess einfacher gestalten. Model Driven Architecture (MDA) und Unified Modeling Language (UML) gelten bei Marktanalysten wie Gartner als Schlüsseltechnologien, die Eigenentwicklungen Zukunftssicherheit und Wiederverwendbarkeit einhauchen. Den Schwung will die in Nürnberg ansässige MID GmbH dafür nutzen, künftig stark zu wachsen. Das unterstrich am Rande der diesjährigen Anwenderkonferenz Insight’07 der geschäftsführende Gesellschafter Wolfgang Dietrich. Im neuen Jahr strebt man ein Plus von 20 % an. Das bislang vornehmlich in hiesigen Gefilden tätige Unternehmen will zudem die internationale Präsenz verstärken. Zukäufe kleinerer Beratungshäuser oder Tool-Anbieter schließt Dietrich explizit ein.

MID ist im MDA-Markt mit der Modellierungsplattform Innovator aktiv, die UML zur durchgängigen Notation vom Geschäftsprozessmodell bis zum Systemdesignmodell nutzt. Das Werkzeug kann die Domain Specific Language (DSL) in Form von UML2-Profilen darstellen. Auf diese Weise wird die Universalität der UML auf das fachliche Konzept eines ausgewählten Anwendungsbereichs zugeschnitten. Ein weiterer Stellhebel, den Weg vom Geschäftsprozess zum servicebasierten Entwurf oder Programmcode zu beschleunigen, ist die MID-eigene Modellierungs-Methodik M3. Sie beschreibt systematisch die einzelnen Prozessschritte, Modellierungsebenen und Rollen. Laut Dietrich gibt die Modellierungsmethodik einschließlich Vorgehensmodell häufig den Ausschlag, Innovator einzusetzen.

Rund 300 „aktive“ Anwenderfirmen mit insgesamt rund 10 000 Nutzern modellieren mit Innovator. Vertreter der BA oder Deutschen Telekom referierten auf der Nürnberger Veranstaltung, wie sie die Plattform dafür nutzen, auf eine serviceorientierte Architektur (SOA) hinzuarbeiten. Die modellhafte Beschreibung der Geschäftsprozesse bis zur Anwendungsarchitektur und IT-Service stellen spezifische Sichten der im Innovator-Repository verwalteten jeweiligen Modelle dar. Über Mapping-Techniken lassen sich fachliche Prozessmodelle aus Aris von IDS Scheer einbinden. Seit Kurzem ist auch der Austausch von Metadaten mit Centrasite der Software AG möglich. Durch dieses Integrationsszenario können sämtliche Informationen einer SOA aus der Design-Umgebung, also Innovator, in die SOA-Registry/Repository-Umgebung wechselseitig zur Verfügung gestellt werden. *Achim Born*

TDM: Transaktionsmanagement mit GoldenGate

Die TDM-Plattform (Transactional Data Management) GoldenGate des gleichnamigen Anbieters ist in der Version 9.5 verfügbar. Mit den Lösungen können Unternehmen ihre Transaktionsdaten in Echtzeit und über verschiedene IT-Umgebungen hinweg erfassen, weiterleiten, transformieren, bereitstellen und überprüfen. Die Software besteht aus Einzelkomponenten (Capture, Trail Files, Delivery), die ihre Aufgaben unabhängig voneinander durchführen und zu verschiedenen Hochverfügbarkeits- und Datenintegrationslösungen zusammengefügt werden können. Die neue Version ist um die Unterstützung von Unternehmensanwendungen wie Oracles E-Business Suite, Peoplesofts Enterprise, Siebel-, JD-Edwards- sowie SAP-Produkte erweitert worden. Auch soll die Replikation von nicht datengebundenen Objekten möglich sein, eine Erleichterung für die Datenbankverwaltung, da sich etwa Anwender einfacher hinzufügen oder gespeicherte Vorgänge auf die Zieldatenbank übertragen lassen. *Susanne Franke*

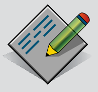
SAP ERP Upgrade: Technik zuerst

Über die Höhe der Wartungskosten möchte SAP ihre Kunden dazu bewegen, endlich auf die jüngsten ERP-Versionen umzusteigen. Gemäß Release-Plan hat für R/3 4.6B und frühere Versionen bereits die Zeit der kundenspezifischen Wartung begonnen. R/3 4.6C befindet sich seit 2007 ebenfalls in der erweiterten Wartung und wird ab 2010 nur noch kundenspezifisch unterstützt.

Aufgrund der Support-Politik des Walldorfer Softwarehauses schien für den deutschen Markt eine Upgrade-Welle in den Jahren 2007/2008 wahrscheinlich. Nach Prognose des Marktanalyse- und Beratungsunternehmens PAC wird diese Welle ihren Höhepunkt allerdings erst 2009 erreichen – sowohl beim Marktvolumen wie auch bei der Anzahl der Projekte. Einen entscheidenden Einfluss hat dabei die Entwicklung der verschiedenen Upgrade-Arten. Zurzeit führen die Unternehmen laut PAC vornehmlich technische Upgrades durch. Die Zahl der funktionalen und strategischen „Aufstiege“ wird erst 2009 die der technischen überschreiten. Die Anwender erneuern zunächst einmal die systemtechnische Basis ihrer Anwendungssoftware und führen erst in Folgeprojekten neue Funktionen oder Geschäftsprozesse ein. Letztgenannte Arbeiten sind deutlich aufwendiger und damit kostspieliger. Das von den PAC-Analysten prognostizierte Verhältnis von Projektanzahl zu -ausgaben belegt den beschriebenen Ablauf.

SAP ERP 6.0: Ramp-Up | Mainstream-Wartung | erw. Wart. (+ 2 %) | erw. Wartung (+ 4 %) | kundenspezifische Wartung
mySAP ERP 2004: Mainstream-Wartung | erw. Wart. (+ 2 %) | erw. Wartung (+ 4 %) | kundenspezifische Wartung
R/3 Enterprise 47x200: Mainstream-Wartung | erw. Wart. (+ 2 %) | erw. Wartung (+ 4 %) | kundenspezifische Wartung
R/3 Enterprise 47x110: Mainstream-Wartung | erw. Wart. (+ 2 %) | erw. Wartung (+ 4 %) | kundenspezifische Wartung
R/3 4.6C: Mainstream-Wartung | erw. Wart. (+ 2 %) | erw. Wartung (+ 4 %) | kundenspezifische Wartung
R/3 3.1I - 4.6B: erw. Wartung (+ 4 %) | kundenspezifische Wartung
Okt. Juni Dez. Dez. März Dez. März März März März März März
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016

KURZ NOTIERT



Onlinedienste: Ab Dezember ist Sage hierzulande im SaaS-Geschäft (Software as a Service) aktiv. Für 21 Euro pro Monat gibt es SageCRM im Mietbetrieb. Die webbasierende CRM-Lösung verfügt über Funktionen zum Erstellen von Umsatzprognosen und Berichten, zum Kampagnenmanagement und zum Steuern von Werbemaßnahmen. Sie lässt sich mit Outlook synchronisieren und in die Office Line von Sage integrieren.

Kombipack: Unter dem Namen PM 10 kündigt Infor seine erste integrierte Performance-Management-Suite (PM) an. Sie vereint die Anwendungen der übernommenen Firmen Extensity MPC und die Analyse-Funktionen von Systems Union MIS. PM 10 umfasst das strategische Management, Planung und Budgetierung, Prozesse zur Rechnungslegung, Finanzkonsolidierung sowie Prognosen. Einige Komponenten der Suite, die unter Windows und Office läuft, lassen sich separat einsetzen.

Aufgepeppt: Das Berliner Softwarehaus Inubit hat die Release 5.0 ihrer gleichnamigen BPM-Suite freigegeben. Neu eingebaut ist beispielsweise ein Portalserver, der dem Portlet-Standard JSR168 folgt. Der Server stellt Monitoring-Ergebnisse, etwa Kennzahlen (KPIs), Drill-Down-Reports und Soll-Ist-Vergleiche in Portlets dar. Die fachliche Modellierung der Geschäftsprozesse lässt sich nun BPMN-konform (Business Process Modeling Notation) erledigen.

Partnerschaft: Die Software AG schloss eine OEM-Lizenzvereinbarung mit Cognos. Sie sieht vor, dass das Darmstädter Unternehmen Cognos 8 BI als Berichts- und Analyse-Standard in die künftigen Versionen seiner Webmethods-Suite integriert. IBMs Absicht, Cognos zu übernehmen, hat angeblich keine Auswirkungen auf den Deal.

Ausgebaut: SER erweiterte seine ECM-Suite (Enterprise Content Management) um ein Archivierungsmodul. Über einen Admin-Client unterstützt der Doxis Filemanager die regelbasierte Archivierung sowohl aus dem Dateisystem als auch aus Microsofts Exchange und Sharepoint heraus. Das Verlagern der Dateien auf günstigen Speicherplatz lässt sich so automatisieren. Der Verweis auf die verschobenen Daten bleibt im Dateisystem sichtbar.

Mehr Cobol: Nachdem Micro Focus kürzlich sein Cobol-Entwicklungssystem für Windows – Net Express – für Eclipse freigegeben hat, ist nun auch die Server Express für die freie Werkzeugplattform erhältlich. Damit lassen sich Cobol-Anwendungen auf Unix- und Linux-Systemen erstellen. Im nächsten Jahr will Micro Focus auch die Mainframe Express Enterprise Edition in den Eclipse-Reigen einbringen.

Winterzeit bei Salesforce

Mit der Winter-08-Release gab Salesforce.com die 24. Generation ihrer On-demand-CRM-Software frei. Das Angebot enthält nun auch die Entwicklungsumgebungen Force.com und Visualforce. Force.com ist ein gehosteter Dienst (Platform-as-a-Service, PaaS), auf dessen Basis ein Entwickler Anwendungen „on-demand“ erstellen können soll. Und mit Visualforce peppt er das Design der Applikationen via HTML, Ajax und Flex auf. Neben den bislang vier etablierten CRM-Programmen bringt die neue Version zwei zusätzliche, Web-2.0-affine Anwendungen mit: Salesforce Content bietet Funktionen wie Tagging (Verschlagwortung), Subscription (Abonnement) und Recommendations (Kommmentierung Dritter), zu dem Zweck, Dokumente und unstrukturierte Daten online zu bearbeiten.

SoftM übernimmt KTW-Teile

SoftM hat vom Masseverwalter der insolventen KTW den Zuschlag für den Erwerb der Vermögenswerte im Geschäftsbereich Semiramis erhalten. Durch diesen Schritt wollen die Münchener den Support für Unternehmen gewährleisten, die ihre ERP-Software über KTW bezogen haben. Am bisherigen Standort der Firma in Kirchbichl will SoftM die Tochtergesellschaft SoftM Solutions GmbH gründen, die einen nennenswerten Teil der KTW-Mitarbeiter weiterbeschäftigen soll.

KTW hatte kurz nach der Münchner Systems Insolvenz angemeldet. Als Grund nennt Chef Reinhold Karner „unerfüllbare Nachforderungen“ der Hausbank BTV als Reaktion auf Lösungsvorschläge. Deren Ausarbeitung war erforderlich geworden, da KTW als Gläubigerin der im vergangenen Oktober pleite gegangenen Semiramis Software mit Forderungsausfällen in zweistelliger Millionenhöhe zu kämpfen hatte. Ähnlich wie im Dezember 2006, als man die Rechte an Semiramis erwarb, sprang SoftM wieder ein, das ERP-Geschäft der KTW zu retten.

HP Universe: Automatisiertes Geschäft

Zusammenfassung

Jürgen Diercks

Auf seiner Benutzerveranstaltung in Barcelona stellte HP eine neue Systemmanagement-Suite vor. Mit der sollen Unternehmen ihre IT-Prozesse weitgehend wartungsfrei steuern können.

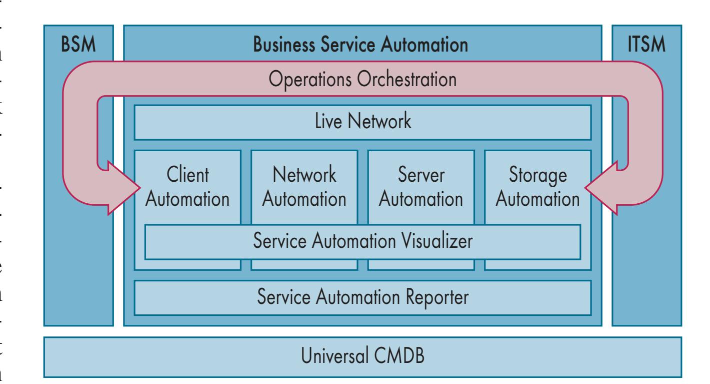
Über 4000 Teilnehmer, 80 Aussteller, 150 Tracks im futuristischem Ambiente des Convention Centers in Barcelona ergaben gebündelt die HP Universe. In den Keynotes sparten die Redner nicht mit den üblichen Lobhudeleien auf die eigene Firma. Und ein lustiger Professor der theoretischen Physik namens Michio Kaku zeichnete ein optimistisches Bild der Zukunft unter dem Motto: „How to compete in the Hyper-Connected World of 2020“. Um hier erfolgreich zu sein, müsse man zunächst mal den „perfekten Kapitalismus“ einführen, in dem der Konsument rundum informiert durchs Leben läuft und die Unternehmen ihm dazu passende, hochwertige, innovative und effiziente Software auf den Leib schneidern. Überhaupt ließe sich demnächst jedes Problem mit entsprechender Technik lösen. Der Mann muss es wissen, denn immerhin zählt ihn das New York Magazine zu den 100 smartesten Insassen der Stadt.

Es gibt jedoch auch Handfestes zu vermelden: Laut Thomas E. Hogan, Senior Vice President von HP, wächst die Softwaresparte des Hauses am schnellsten und stellt mittlerweile das profitabelste Segment des nach eigenen Aussagen sechstgrößten Softwareanbieters. Im fiskalischen Jahr 2007 setzte HP 2,33 Mill. US-Dollar um, ein Zuwachs um 79 %. Der Gewinn stieg sogar um 306 %. Mittlerweile ernährt die Software 7000 Angestellte bei HP.

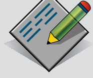
Wie Hogan weiter ausführte, unterstützt die IT das Unternehmen bislang nur unzureichend bei der Koordination der zahlreich anwesenden Tools, Teams und Prozesse. Er prognostiziert sogar eine regelrechte Explosion der IT-Komplexität. Beispielsweise würden bereits im Jahr 2009 auf einen echten vier virtuelle Server kommen. Als Schutz vor dem drohenden Chaos kündigte er die Produkt-Suite „Automated Operations 1.0“ an. Sie soll den Lebenszyklus aller Geschäftskomponenten auf einem gemeinsamen Datenmodell (Universal CMDB) sowie einem integriertem Workflow Layer abbilden und somit helfen, den gesamten IT-Betrieb zu automatisieren. Die Suite fasst HPs bekannte Systemmanagement-Angebote mit den von Opsware übernommenen Produkten zusammen.

Automated Operations besteht aus: IT Service Management (ITSM, ITIL-Unterstützung), HP Business Service Automation (BSA), Bereitstellung von IT-Resourcen (Provisioning) sowie Business Service Management (Management von Geschäftsprozessen, Netzen, Infrastruktur, Anwendungen). Über die Orchestrierungsfunktionen lassen sich alle Komponenten miteinander verknüpfen (siehe Grafik).

In den SaaS-Markt (Software as a Service) will HP groß einsteigen. Angeblich wächst dieser Bereich drei- bis viermal schneller (20 % pro Jahr) als das ERP-Segment. HP verspricht seinen Kunden, dass sie über dieses Modell bis zu 20 % der Kosten für IT-Management sparen können. (jd)



KURZ NOTIERT



Automatisiert: Microsoft hat das V-Modell XT in Visual Studio Team System integriert. Ein Prozesstemplate-Generator soll die Vorgaben des Vorgehensmodells im Team Foundation Server abbilden. Letzterer ist die zentralen Komponente der Entwicklungsplattform. Interessierte können sich den Generator von www.codeplex.com/VModellXTTFS kostenlos herunterladen.

Studiotermin: Zend hat die Beta-Version seiner PHP IDE „Studio for Eclipse“ veröffentlicht. Sie basiert auf den Studio-Produkten des Herstellers und dem Eclipse PHP Developers Tools (PDT) Project. Die IDE richtet sich an professionelle PHP-Entwickler, die eine Umgebung für die Erstellung webbasierter Geschäftsanwendungen benötigen www.zend.com/en/products/studio/downloads.

Durchgängig: Vom Modell zum Einsatz

Telelogic hat seinen System Architect, ein Werkzeug für die Modellierung von Unternehmensarchitekturen, in der Version 11.0 mit Tau 4.0 verknüpft. Tau ist ein Programm für die modellgetriebene Anwendungsentwicklung. Anwender sollen damit einen modellbasierten Workflow erstellen können – von der Unternehmensarchitektur und dem Geschäftsmodell bis zur Implementierung, und zwar im Zusammenspiel mit der Anforderungsmanagementlösung DOORS. Beide Umgebungen teilen sich ein Repository.

Laut Telelogic haben die Benutzer die Möglichkeit, mit dem System Architect entworfene Geschäftsprozesse in UML-Modelle umzusetzen, die wiederum zur Steuerung der Systementwicklung dienen. Umgekehrt lassen sich in Tau durchgeführte Änderungen in System Architect auf ihre Eignung prüfen. Die Visualisierung der Modelle soll sich nach der Rolle des Anwenders richten. Die Werkzeugkombination unterstützt die gängigen Modellierungsstandards (BPEL, BPMN, UML). Nach Einschätzung des IDC-Analysten Steve Hendrick lässt sich mit der Umgebung eine SOA aufsetzen, von den Anforderungen bis hinunter zu den Webservices.

Susanne Franke

Microsofts Roadmap für Composite Applications

Unter dem Codenamen Oslo will Microsoft SOA-Techniken entwickeln, die Eingang finden sollen in den Biztalk Server 6, Visual Studio 10, System Center 5 sowie das .Net Framework 4. Sogenannte Composite Applications auf Oslo-Basis kann der Kunde sowohl lokal (On-Premise) als auch über ein Mietmodell betreiben. Entwickelt wird künftig modellgetrieben, das Modellieren mit Visual Studio sollen Geschäftsanalysten, Architekten und Entwickler gemeinsam erledigen. Motto: „Das Modell ist die Anwendung.“

Oslo soll die einzelnen Modelle in ein Gesamtmodell überführen, in dem jeder Beteiligte stets alle Einzelteile sehen kann. Tools wie das Design-Werkzeug Whitehorse für die Erzeugung und Implementierung von Systemen in einer SOA will Microsoft weiterentwickeln. Einen „universellen“ Editor hat der Konzern bereits vorgestellt. Er soll End-to-End-Ansichten von Modellen für Geschäfts- oder IT-Prozesse aller Art darstellen können.

Die Roadmap sieht auch ein neues Repository für Metadaten, Biztalk-Prozesse, Webservices sowie Code-Artefakte aus Visual Studio vor. Ein Internet Service Bus (Weiterentwicklung der Biztalk Services) soll Prozessdienste leisten. Der ISB hat grundsätzlich die gleichen Aufgaben – etwa Nachrichten-Routing und Datenwandlung – wie ein herkömmlicher Enterprise Service Bus. Allerdings nutzt der ISB Internetprotokolle und arbeitet Firewall- sowie organisations- und plattformübergreifend. *Susanne Franke*

Verbindung via Performance Point Server

Mit dem Office Performance Point Server 2007 stellt Microsoft ein erstes eigenes Angebot für das Corporate Performance Management vor. Es verbindet Analyse-, Planungs- und Budgetierungsfunktionen mit einem erweiterten Excel-Frontend. Die Software setzt für OLAP und Data Mining die SQL Server Analysis Services 2005 voraus.

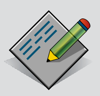
Der integrierte Planning Business Modeler hilft beim Erstellen von Planungs- und Geschäftsmodellen. Einige Modelle mit vorgefertigten Dimensionen und Geschäftsabläufen stehen zur Verfügung. Als Eingabeoberfläche dient ein neues Excel-Plug-in. Schließlich gehört zur Plattform der Business Scorecard Manager 2005, dessen Technik vom aufgekauften Business-Intelligence-Anbieter ProClarity stammt. Damit sollen Anwender Kennzahlen, miteinander verknüpfte Scorecards sowie Strategy Maps erzeugen können. Unternehmensweite Modelle für Scorecards, Analysen und Planung lassen sich laut Microsoft auch auf Abteilungsebene für das Performance Management heranziehen. So sollen Kunden eine klare Sicht auf organisationsweite Leistungsparameter erhalten. Mit der ERP-Software Dynamics arbeitet der Server zusammen. Der Konzern hofft auf reges Interesse der Partner, die Dynamics und den SQL Server im Angebot haben.

Susanne Franke

CRM und ERP für Mittelstand

Speziell mittelständische Unternehmen will die Anfang 2007 in Wien gegründete Kinamu AG mit CRM- und ERP-Software beliefern. Bei der Firma, die gerade eine deutsche Tochter in München aufbaut, handelt es sich um ein Spin-off ehemaliger SAP-Manager. Geschäftsidee ist, den Kunden solide Technik zum kleinen Preis anzubieten und das ohne Fixkosten und Risiko. Kinamu will Komplettpakete aus Hardware, Software, Betrieb, Support und definierten Service Levels anbieten. Die Mittelständler können ihre monatlichen Kosten laut Anbieter klar kalkulieren und müssen kein internes Spezialwissen aufbauen, da die Lösungen remote laufen. Wenn möglich, basieren die Anwendungen auf Open-Source-Produkten. So vertreibt Kinamu etwa SugarCRM. Im ERP-Umfeld ist SAP ERP 6.0 Kernkomponente. In den Paketen sind diese beiden Applikationen miteinander verzahnt.

KURZ NOTIERT



Version 2: Für High-End-PCs und Server-Storage-Lösungen hat Samsung Electronics seine 64-GByte-SSDs (Solid State Drives) in den Formfaktoren 1,8 und 2,5 Zoll mit SATA-II-Interface bestückt. Die 64 jeweils 8 GBit großen 50-nm-SLC-Flash-Chips (Single-Level-Cell) erreichen etwa 120 MByte/s lesend und 100 MByte/s schreibend bei einer Leistungsaufnahme von weniger als einem Watt.

Mal 10: Reldata hat mit dem Virtualisierungs-Gateway 9240 seine Produktpalette um einen 10-GBit/s-fähigen SAN/NAS-Gateway erweitert. Zu den sechs GE- und je zwei U320-SCSI- FC-Ports gesellen sich nun optional ein 10-GE-Target-Port sowie ein 3-GBit/s-SAS-Interface.

Mehr Platz: Der im November von Dell übernommene Speicherhersteller Equallogic hat seine Produktpalette mit dem Disk-Array PS3700X iSCSI erweitert. Es ist mit den neuen, 400 GByte fassenden SAS-Festplatten der Cheetah-NS-Serie von Seagate bestückt und kommt damit auf eine Gesamtkapazität von 6,4 TByte brutto.

Mit Alternative: Hitachi hat QLogics iSCSI-to-FC-Router SANbox 6140 in seine Midrange-Storage-Familie Adaptable Modular Storage und Low-End-Modelle des Workgroup Modular Storage integriert. Dadurch sind sie künftig auch über iSCSI ansprechbar.

Zum Anfassen: Die Münchener Hetec Datensysteme GmbH hat ihre Screensplitting-KVM-Switches mit Touchscreen-Unterstützung ausgestattet. Über den V-Switch soll der Touchscreen als Eingabegerät Tastatur und Maus ersetzen können.

IBMs BlueGene-Systeme dominieren Green500

Blaugrün

Nikolai Zotow, Susanne Nolte

Auf den ersten Blick wirkt die neue Green500-Liste eintönig. Denn die ersten 27 Plätze belegen fast ausschließlich BlueGene-Systeme von IBM.

Die vorderen 5 Plätze haben sich die neuen BlueGene/P-Systeme gesichert, die Plätze 7 bis 27 die älteren L-Systeme – und 5 der 26 Installationen stehen bei IBM selbst. Fast verloren wirkt dazwischen der Bio-X2 – ein PowerEdge-1950-Cluster an der Stanford University – auf Platz 6.

Nunmehr zum dritten Mal sortiert die Green500 (www.green500.org) die Supercomputer der frisch veröffentlichten Top500 (siehe Seite 14 in diesem Heft) nach ihrer Energieeffizienz. Das Ranking erfolgt nach Mflops pro Watt, wobei – wenn vorhanden – die gemessenen Werte gelten, andernfalls die Peak-Werte des Stromverbrauchs. Andere Kenngrößen finden hier keine Berücksichtigung.

Spitzenreiter ist der in der Top500 auf Platz 121 gelistete Blue Gene/P des britischen Science and Technology Facilities Council mit 357,23 MFlops/W, gefolgt vom „Genius“ der Max-Planck-Gesellschaft – Deutschlands viertgrößtem Rechner – mit 352,25 MFlops/W (Top500: Platz 40). Hinter dem IBM-eigenen System in Rochester (346,95 MFlops/W; Top500: Platz 24) belegt die momentan leistungsfähigste zivil genutzte Installation der Welt, JUGENE am Forschungszentrum Jülich, mit 336,21 MFlops/W Platz vier. Mit etwas Abstand auf den fünften BlueGene/P (310,93 MFlops/W) folgt das oben genannte Dell-System Bio-X2 mit 245,27 MFlops/W. Die auf den Plätzen 7 bis 27 geführten Großrechner differieren kaum in den Effizienzwerten (210,56 bis 203,77 MFlops/W). Darunter befinden sich auch der zweite Jülich-Rechner JUBL auf Platz 14 (Top500: Platz 28) und auf dem 22. Platz mit dem bestplatzierten Top500-Rechner BlueGene/L am LLNL der erste militärisch genutzte Supercomputer der „grünen“ Liste.

Lange suchen muss man in der Green500, bis man andere deutsche Aushängeschilder findet: Der in der Top500 auf Platz 15 vermerkte HLRB-II (SGI Altix 4700) des Leibniz-Rechenzentrums belegt hier mit ganzen 22,00 MFlops/W den 453. Platz. Knapp darüber liegt auf Platz 451 eine weitere SGI Altix 4700, diesmal die der TU Dresden mit 22,32 MFlops/W (Top500: Platz 112). Etwas besser ist der HP-Opteron-Cluster der Uni Karlsruhe platziert: Mit seinen 24,32 MFlops/W hat er gerade einmal Platz 444 in der Green500 erreicht (Top500: Platz 114). Zumindest der in Karlsruhe frisch installierte Xeon-Cluster nimmt schon jetzt mit 142,21 MFlops/W Platz 36 der Green500 und Platz 104 der Top500 ein.

Der schlechteste Energieverwerter der in deutschen Landen beheimateten Top500-Systeme ist aber der NEC SX8/576M72 am HWW der Universität Stuttgart auf Platz 496 (6,2 MFlops/W). Ihn untertreffen nur noch der 2002 gebaute Earth-Simulator (5,6 MFlops/W), die beiden US-amerikanischen Itanium-Cluster Teragrid (5,08 MFlops/W) und Thunder (4,06 MFlops/W), gefolgt vom Schlusslicht ASCI Q (AlphaServer SC45; 3,65 MFlops/W) in Los Alamos.

Aktive Lichtwellenleiter für 10 GE

Auf der Supercomputing Conference November 2007 hat Finisar erste aktive Lichtwellenleiter für 10-Gigabit-Ethernet vorgestellt. Sie besitzen in den Steckern Transceiver, die die Signale zwischen den Kupfer-Ports der Switches und HBAs auf der einen und dem optischen Kabel auf der anderen Seite konvertieren. Finisar liefert die Laserwire genannten Kabel konfektioniert in 3, 5, 10 und 30 Meter Länge und als Sonderanfertigungen. Die Konnektoren sind dabei nicht breiter als die von RJ-45 oder SFP (Small Formfactor Pluggable).

Neben 10GE unterstützen die Kabel Fibre Channel mit 1 bis 8 GBit/s, externe SAS/SATA-Verbindungen, externes PCIe, HDMI (High Definition Multimedia Interface), Displayport sowie proprietäre High Speed Interconnects. Für Infiniband würde nur die – theoretisch existierende – 10-GBit/s-Variante QDR x1 (Quad Data Rate auf einem 2,5-GBit/s-Kanal) infrage kommen, für die es aber weder Switches noch HBAs gibt. Intel hatte bereits im Sommer auf der ISC 2007 in Dresden solche Active Cables für die herkömmlichen elektrischen CX4-Infiniband-Anschlüsse mit 10 (SDR x4) und 20 GBit/s(DDR x4) vorgestellt.

Gegenüber den etwa bei Infiniband und SAS verwendeten Kupferkabeln kann Glasfaser vor allem durch das wesentlich geringere Gewicht, den geringeren Biegeradius, die kleineren Stecker für eine höhere Port-Dichte sowie einen geringeren Stromverbrauch punkten – Finisar gibt eine Leistungsaufnahme von typischerweise weniger als 0,5 W pro Ende an. Zum Vergleich: Die PHYs der – noch nicht spezifizierten – Twisted-Pair-Kabel für 10GBase-T saugen mehr als 10 Watt aus dem Netzteil. Allerdings werden bei 10GE und Fibre Channel schon jetzt vornehmlich Lichtwellenleiter verwenden. Hier argumentiert Finisar, dass die Glasfaser-Ports sich beim 10GE bisher nur in den Kernnetzen durchsetzen konnten, da sie für Server-Anbindung noch zu teuer sind. Das will der Hersteller durch entsprechende Kupferports für Motherboards und HBAs ändern.

SDK verbessert Projektmanagement

Als Teil der neuen Version seines SCM-Systems Perforce Server 2007.3 hat Perforce ein SDK für das Defect Tracking Gateway angekündigt. Das Software Development Kit soll Kunden und Händler in die Lage versetzen, in kurzer Zeit Plug-ins zu entwickeln, die ihre jeweilige ALM-Lösung (Application Lifecycle Management) an individuelle Anforderungen anpassen. Über eine eingebaute Replication Engine reicht das Defect Tracking Gateway vom Anwender vorgenommene Änderungen weiter.

Die Preise für den Perforce Server 2007.3 beginnen bei 800 US-Dollar für eine Lizenz. Eine kostenlose 45-Tage-Testversion bietet der Hersteller über seinen Website ebenfalls an (www.perforce.com) – inklusive technischen Supports.

3D-Dokumentation für die Nachwelt

Kein Kulturgut bleibt ewig erhalten. Sie erodieren im Laufe der Jahrhunderte oder stehen Bauprojekten im Wege. Letzteres betrifft beispielsweise Felszeichnungen am Karakorum Highway, von denen ein großer Teil bis 2016 durch die Aufstauung des oberen Indus zerstört werden soll. Eine Forschungsgruppe der Heidelberger Akademie der Wissenschaften will mithilfe von 3D-Laserscannern und Compterfräsen versuchen, detailgetreue Repliken zu erstellen und so zumindest einen Teil der Gravuren zu erhalten.



Quelle: Akademie

Eine weitere Forschungsstelle plant, mit computergestützten 3D-Modellen und Animationen der Weltöffentlichkeit in Stein gemeißelte heilige Texte aus buddhistischen Sutren zu präsentieren. Die datenbankbasierte Dokumentation über die Inschriftenorte in der chinesischen Provinz Shandong soll zukünftig über das Internet zugänglich sein. Weitere Informationen unter www.haw.baden-wuerttemberg.de.

3D-Scans mit 64-Bit-Leistung

Mit den 64-Bit-Versionen ihrer Scan-Software umgeht Inus Technology, Inc. (www.rapidform.com) die Beschränkung auf 2 GByte Arbeitsspeicher und kann jetzt bis zu 128 GByte RAM nutzen.

Rapidform XOS (Scandatenverarbeitung) und XOR (Redesign) bieten unter anderem Punktwolkenverarbeitung und erzeugen editierbare parametrische Volumenmodelle, die in CAD-Systemen unmittelbar weiterverarbeitet werden können. Die 64-Bit-Ausgabe des für die Qualitätsprüfung entwickelten Rapidform XOV (Verifier) hat das koreanische Unternehmen für Anfang 2008 in Aussicht gestellt. Voraussetzung für den Betrieb von Rapidform XOS 64, XOR 64 und XOV 64 sind ein Rechner mit 64-Bit-Prozessor sowie Windows XP x64 oder Vista x64.

Repository-Tool für Build-Werkzeuge

Das Maven Archiva Team hat die finale Version 1.0 des Repository Manager Archiva angekündigt (maven.apache.org/archiva). Archiva arbeitet mit Build-Werkzeugen wie Maven, Continuum und Ant zusammen. Es erlaubt unter anderem das Durchsuchen und Sichern von Repositories und identifiziert unbekannte Artefakte. Darüber hinaus kann es als Proxy Cache für andere globale Repositories fungieren.

Nächste Generation des VMware Server

Es gibt bereits die Beta-Version des VMware Server 2 zum freien Download. Mit der neuen Version sollen sich die Bedienung und das Installieren virtueller Maschinen (VM) dank intuitiver Management-Schnittstelle vereinfacht haben. VMware nutzt ein Web-Interface, sodass Linux- wie Windows-Benutzer eine einheitliche Benutzer-Oberfläche vorfinden. Außerdem haben die Entwickler das Hardwarespektrum erweitert und unterstützen über 30 Varianten von Gastbetriebssystemen, unter anderem neue Linux-Distributionen und Windows Server 2008. Das Virtual Machine Interface (VMI) erweitert die Kommunikation zwischen den VMs und der Virtualisierungsebene. Hinzugekommen ist außerdem die Nutzung von USB 2,0, bis zu 8 GByte RAM pro VM, maximal zweier virtueller SMP-Prozessoren und von 64-Bit-Gastsystemen (www.vmware.com/go/server2_beta).

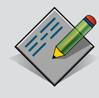
SPECjms2007 misst Leistung von MOM

Ein Benchmark, der die Leistung von „Message-oriented Middleware“ (MOM) misst, ergänzt neuerdings das Portfolio der Standard Performance Evaluation Corporation (SPEC). Ende Oktober hat sie den neuen SPECjms2007 veröffentlicht. MOM koordiniert die Kommunikation von Anwendungen untereinander und dient als Basis für „Service Oriented Architecture“ (SOA) oder „Event Driven Architecture“ (EDA).

Den zugrunde liegenden Standard „Java Message Service“ (JMS) unterstützen die meisten namhaften Softwarehersteller, neben allen J2EE-Produkten gilt dies beispielsweise für Tibcos Enterprise Message Service, IBMs Websphere MQ oder Oracles Advanced Queueing. SPECjms2007 ermöglicht sowohl den Vergleich der Leistungsfähigkeit solcher Produkte auf Basis von Standardszenarien als auch die Analyse von MOM-Lösungen unter benutzerspezifischen Bedingungen. Zur Bestimmung der Systemleistung bildet SPECjms2007 den Nachrichtenfluss in der Supply Chain einer Supermarktkette ab.

SPECjms2007 entstand in einem Zeitraum von zwei Jahren als gemeinsames Projekt der TU-Darmstadt sowie der Firmen IBM, Sun, Oracle, Bea, Sybase und Apache unter Federführung des SPEC-Java-Komitees. Ergebnisse liegen bei der SPEC (www.spec.org) noch nicht vor. Der SPECjms2007 kostet 1800 US-$, für Non-Profit-Nutzer 450.

KURZ NOTIERT



ACML 4.0: Ein Update des Linear Algebra Package (LAPACK) sowie eine Optimierung für die neuen Quad-Core-Opteron, bezogen auf Multithreading und Streaming Extensions (SSE), zählen zu den herausragenden neuen Features, mit denen AMD seine frei erhältliche Math Core Library (ACML) in der Version 4.0 ausgestattet hat (developer.amd.com).

Grafisch rechnen: Unterstützung für die 64-Bit-Version von Windows XP samt neuer Source-Codes zum Durchführen von Berechnungen enthält CUDA 1.1, die neue Version der Compiler für die GPUs auf Nvidias Grafikkarten. Außerdem bieten die neuen Display-Treiber die Unterstützung für CUDA, sodass das zusätzliche Installieren von Software entfällt (www.nvidia.de/cuda).

Im Paket: Mit der Premium Edition von Parallels Desktop 3.0 bietet SWsoft ein Paket zur Virtualisierung auf Intel-Macs mit Werkzeugen von Acronis und Kaspersky zur Sicherung von Windows-Gastsystemen. Erhältlich ist das Bundle im neuen deutschen Onlineshop von SWsoft (www.parallels.com/de/products/desktop/).

KVM-over-IP von Raritan

Bei seinem neuen Top-Modell hat Raritan die Ausstattung erweitert. Der KVM-over-IP-Switch aus der Serie Dominion KX II ist jetzt mit 64 Ports ausgestattet. Der DKX2-464 besitzt zwei redundante Netzteile und zwei Gigabit-Ethernet-Anschlüsse. Die bis dato lästige Anpassung der Maus an die ferngesteuerte grafische Oberfläche soll durch Absolute Mouse Synchronization automatisch von der Hand gehen. Virtual Media, die Einbindung von Laufwerken vor Ort an entfernte Systeme, erleichtert Installation und Wartungsaufgaben. Mit einer Auflösung von 1600 × 1200 bietet das Gerät eine angenehm große Arbeitsfläche für Administratoren. Der Preis für das Dominion KX2-464 beträgt 7476 Euro.



Alles dabei: Nicht nur 64 Ports, sondern auch PS/2 und USB nebst doppeltem Netzteil und Gigabit-Ethernet kennzeichnen den KVM-over-IP Switch von Raritan.

LogMeIn hilft beim Backup und steuert Macs

Inkrementelles Backup auf dedizierte PC im Firmennetz sowie die Option, den Prozess anhalten und wieder aufnehmen zu können, bietet LogMeIn mit seiner neuen Version von Remote Backup an. Der Dienst kostet je Rechner 39,95 US-$ pro Jahr, bei mehreren sinkt der Preis auf 34,95 US-$ (www.logmeinbackup.com).

Für Liebhaber der Systeme von Apple gibt es neuerdings LogMeIn Free for Mac. Es handelt sich um eine kostenlose Dienstleistung, die Nutzern von Apples Computern unter Mac OS X den Weg eröffnet, von jedem Mac oder PC aus über das Internet auf ihren Rechner zugreifen zu können, wobei der Zugang zum Mac und den dortigen Dokumenten sowie Anwendungen durch Verschlüsselungsmethoden gesichert ist.

Als Beta-Version bietet der Dienstleister „LogMeIn Rescue for Mac“ zum Test. Das Angebot „Software as a Service“ wendet sich an IT-Support-Dienstleister, die damit die Macs ihrer Kunden aus der Ferne diagnostizieren, reparieren oder verwalten können (www.logmein.com).

Virtual Iron 4.2 erweitert Dynamic

Multi-Pathing für virtuelle Server über Fibre Channel, Komplettabzüge von VMs im laufenden Betrieb mit Live-Snapshot und die Möglichkeit, die Größe von Plattengruppen und virtuellen Laufwerken on demand zu vergrößern, sind die herausragenden Features der neuen Version des Xen-basierten Virtualisierers Virtual Iron 4.2. Hinzu kommt außerdem eine umfangreichere Unterstützung von Betriebssystemen einschließlich Red Hat Enterprise Linux 5 (RHEL) und Suse Linux Enterprise Server 10 (SLES). Mit dabei sind Werkzeuge für VMs als ISO-Image, die für den Administrator als virtuelle CD-ROMs erscheinen. Damit kann er sie einfacher als bisher verteilen und Upgrades anbieten.

Die Preise, 799 Euro für Extended und 499 für die Enterprise Edition, bleiben unverändert (www.virtualiron.com). Distributor in Deutschland ist die AVnet Technology Solutions (Mail an: volker.schlenker@avnet.com).

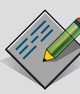
Aufgeholt: Fedora 8 turnusgemäß fertig

Da sich die Entwickler vorgenommen hatten, die rund einmonatige Verspätung der Vorversion Fedora 7 aufzuholen, blieben ihnen diesmal lediglich fünf Monate für die Fertigstellung von Fedora 8 (Codename Werewolf) und das Wiederherstellen des gewohnten halbjährlichen Rhythmus. Wie gewohnt spendierten sie ihrer Distribution reichlich Aktualisierungen, wobei sich über den gesamten Verlauf des Entwicklungszyklus erstmals auch Nicht-Red-Hat-Programmierer ernsthaft in das Open-Source-Projekt einbringen konnten.

Zu den größeren Erweiterungen gehören zum einen Suns Open-Source-Java-Runtime IcedTea sowie der modulare Soundserver PulseAudio, der den ESD-Daemon ersetzt. Der grafische Mixer kann dank virtueller Ausgabekanäle auch mehreren Anwendungen Zugriff auf die Soundausgabe gewähren. Hier kam es bislang immer wieder zu Hakeleien zwischen Applikationen. Dank des aktuellen Kernels 2.6.23 (im Update-Zweig lautet die Versionsnummer inzwischen 2.6.23.3) verfügt Fedora 8 über eine gute Treiberausstattung – sofern Open-Source-Treiber existieren. In Sachen proprietärer oder patentrechtlich zweifelhafter Software gibt sich das Projekt nach wie vor ähnlich restriktiv wie Debian und liefert diese schlicht nicht mit aus.

Mit ATrpms (atrpms.net), FreshRPMs (freshrpms.net), Dribble (dribble.org.uk) oder Livna (rpm.livna.org) gibt es zwar eine Reihe von Repositories, die diese Lücken füllen, aber um deren Integration in die Paketverwaltung via *yum* muss sich der Anwender schon selbst kümmern. Ein Hoffnungsschimmer für Benutzer: Dribble, rpm.livna.org und FreshRPMs arbeiten unter dem Projektnamen RPM Fusion (www.rpmfusion.org) derzeit an einem Zusammenschluss ihrer Softwarearchive. Eigentlich wollte man zum Erscheinen von Fedora 8 fertig sein, inzwischen gilt Januar 2008 als realistischer Termin.

KURZ NOTIERT



Verschoben: KDE 4.0 wird nicht wie ursprünglich geplant Mitte Dezember, sondern erst Mitte Januar endgültig freigegeben. Die Entwickler wollen die zusätzliche Zeit zur Fehlerbehebung nutzen, damit die neue Version den Qualitätsansprüchen des Projekts genügt (www.kde.org).

Freigabe: Daniel Bernstein hat den Quellcode seines Mailservers Qmail in der Version 1.03 als Public Domain freigegeben. Auf der Website des Projekts (www.qmail.org/top.html) ist diese Neuigkeit allerdings noch nicht angekommen.

Bug-Fix-Release: Mit der jetzt veröffentlichten Version 2.3.1 bringt das OpenOffice.org-Projekt seinem Office-Paket keine neuen Fähigkeiten bei. Es dient ausschließlich der Softwarepflege. So behoben die Entwickler fast 50 Fehler, darunter eine Sicherheitslücke, mit der Anwender das integrierte Datenbank-Modul HSQLDB gezielt dazu bewegen können, Java-Code auszuführen (development.openoffice.org/releases/2.3.1.html).

Spezial-Linux: Montavista (mvista.com) will mit Erscheinen dieser Ausgabe sein auf Telekommunikationssysteme spezialisiertes Carrier Grade Linux 5.0 freigeben. Es läuft auf x86-, x86_64 und PowerPC-Systemen, die die PICMG-Standards ATCA und MicroTCA (www.picmg.com/v2internal/specifications.htm) unterstützen, und erfüllt die Carrier-Grade-Spezifikation der Linux Foundation (www.linux-foundation.org).

Fusion: Die Herstellervereinigungen Live (www.linux-verband.de) und Linux Solutions Group (LiSoG, www.lisog.de) haben ihre Gespräche über gemeinsame Projekte ausgeweitet und sprechen über eine potenzielle Fusion der Organisationen.

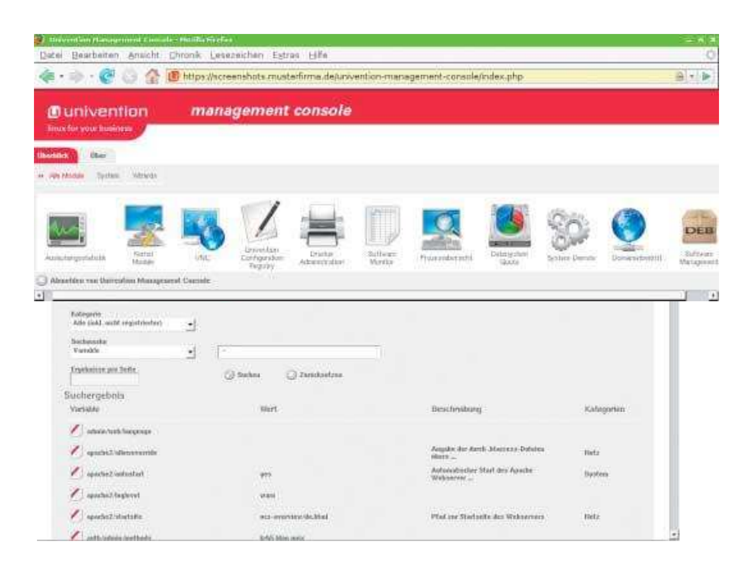
Erstes Update für RHEL5

Für die jetzt verfügbare erste große Aktualisierung von Red Hat Enterprise Linux (RHEL, www.redhat.de/rhel), die Release 5.1, haben sich die Entwickler neben den üblichen Bugfixes, Treiber- und Sicherheitsupdates vor allem die Virtualisierung noch einmal gründlich vorgenommen. So aktualisierten sie Xen auf Version 3.1 und verbesserten die Unterstützung der Hardware-Virtualisierungsfunktionen von AMD- und Itanium-Prozessoren. Die für einen schnellen Betrieb erforderlichen paravirtualisierten Treiber für Linux und Windows will Red Hat separat vertreiben. Da für Red Hat wegen bestehender Zertifizierungen ein Wechsel der Kernel-Version mit sehr viel Aufwand verbunden wäre, integrierten die Entwickler eine Reihe aktueller Hardwaretreiber aus neueren Kernel-Versionen, beispielsweise der Netzwerktreiber *e1000e* für Intels neuere Chipsätze, oder diverse Storage-Controller. Abonnenten des Red Hat Network steht die neue RHEL-Version wie gewohnt kostenfrei zur Verfügung. *iX* wird sich RHEL 5.1 in der nächsten Ausgabe genauer ansehen.

UCS 2.0 fertiggestellt

Mit Erscheinen dieser Ausgabe will Univention die Version 2.0 seines Corporate Server (UCS) für x86-, x86-64- und PowerPC-Systeme (IBM iSeries) freigeben. UCS basiert zwar auf Debian 4.0/Etch, die Bremer aktualisierten aber eine Reihe von Paketen. Vor allem die Integration ins neue UCS-Managementsystem erforderte einige Modifikationen an den Basispaketen. Es bietet eine komfortable Benutzerschnittstelle zum LDAP-basierten Univention Directory Manager, über den domänenweit die Systemverwaltung und -konfiguration erfolgt. Dem KDE-Desktop spendierten die Entwickler einige Extra-Updates, damit der Groupware-Client Kontact Enterprise-Ansprüchen genügt. Via Xen 3.1 lassen sich virtuelle Linux- und Windows-Installationen auf einem UCS-System betreiben. Weitere Details zu den aktualisierten Komponenten finden sich unter www.univention.de/ucs.html.

Bis auf die proprietären Anwendungen von Drittherstellern, beispielsweise Adobes Acrobat, stehen alle einzelnen Komponenten unter der GPL oder einer anderen OSI-konformen Lizenz (www.opensource.org/licenses). Das betrifft auch die Eigenentwicklungen wie das Managementsystem, die Univention im Sommer ebenfalls unter der GPL freigab. Für den Vertrieb der Software verwenden die Bremer ein Support-Subskriptionsmodell; so kostet das Basispaket für einen Server und 10 User/Clients circa 590 € pro Jahr.



20. DOAG-Konferenz

Alles eins

Christian Kirsch



Was als reine Datenbankkonferenz begann, hat sich mittlerweile zur Großveranstaltung für alle Oracle-Produkte entwickelt. In diesem Jahr trafen sich die DOAG-Mitglieder zum 20. Mal, erstmals in Nürnberg.

Noch steht das RDBMS im Mittelpunkt der Deutschen Oracle-Anwendergruppe (DOAG). Andere Produkte aus dem stetig wachsenden Angebot der US-Firma spielen jedoch von Jahr zu Jahr eine größere Rolle. So mussten sich die Teilnehmer diesmal zwischen 200 Vorträgen in 16 parallelen Vortragsreihen entscheiden, die von Neuerungen in Oracle 11g über serviceorientierte Architekturen bis zu Peoplesoft- und Business-Intelligence-Anwendungen reichten.

Im Datenbankteil bildete erwartungsgemäß die neue Version Oracle 11g den Schwerpunkt. Im Vordergrund standen Neuerungen bei Hochverfügbarkeitsfunktionen, im Optimizer, bei der XML-Verarbeitung und der Nutzung von Flashback. Überraschend viel Zulauf hatten Open-Source-Themen. So musste Robert Szilinski seinen Beitrag über den Einsatz von Eclipse, Perl und Subversion ein zweites Mal halten, weil der Raum nicht alle Interessenten hatte fassen können.

Professor Vossen vom ERCIS (European Research Center for Information Systems) in Münster beschäftigte sich in seiner Keynote mit dem Web 2.0 und seinen möglichen Auswirkungen auf die Softwareindustrie. Dabei blieb er jedoch bei der Beschreibung des Status quo stehen. In weiten Teilen erinnerte der Beitrag an Politikeransprachen zu technischen Themen: Ehrfurchtsvoll der neuen, noch nicht richtig verstandenen Technik gegenüber, die nun aber wirklich endlich den großen Durchbruch bringen soll. Die interessante Frage, wer all die Inhalte in Blogs und Wikis lesen, geschweige den, pflegen und aktualisieren soll, ließ der Redner folglich unbeantwortet.

Zufriedenheit mit dem Support wächst

Der Veranstalter DOAG zeigte sich ebenso zufrieden mit dem neuen Tagungsort Nürnberg wie viele Teilnehmer. Dass etwas weniger als die erwarteten 2000 Besucher gekommen waren, schrieb er dem drohenden Lokführerstreik zu. Die Gruppe sieht sich weiterhin als Vertreterin aller Oracle-Kunden, unabhängig vom verwendeten Produkt. Ein Ausdruck dieses Anspruchs sei die Aufnahme eines Vertreters der ehemaligen Siebel-Anwendergruppe in den Vorstand. Mittlerweile hat die DOAG über 3000 Mitglieder und wächst jährlich um rund 10 %.

Als Ergebnis ihrer jährlichen Umfrage konstatierte sie eine gewachsene Zufriedenheit mit dem Oracle-Support. Gegenüber dem Vorjahr hatten mit 502 mehr als doppelt so viele Mitglieder die Fragen beantwortet. Von ihnen waren 48 % „zufrieden“ oder „sehr zufrieden“ – im Vorjahr hatten sich nur 37 % so geäußert. Von der DOAG sehen knapp über 50 % ihre Interessen in diesem Bereich „gut“ oder „sehr gut“ vertreten. Das mag auch daran liegen, dass die Vereinigung inzwischen Mitglied der Internationalen Oracle User Community ist und dadurch direkter Kontakt zum Oracle-Hauptquartier besteht. (ck)

SGI und Oracle virtualisieren

Auf der diesjährigen Oracle World demonstrierte SGI gemeinsam mit SWsoft (www.swsoft.com) die Virtualisierung von Oracle-Datenbanken auf seinen Altix-Servern. Die Linux-Variante von SWsofts Virtuozzo lieferte dabei den Unterbau für einen Oracle-Cluster. Nach Angaben der beiden Unternehmen lassen sich in dieser Umgebung einzelne Oracle-Instanzen ohne Unterbrechungen und Informationsverlust zwischen verschiedenen physischen Servern verschieben. Oracle präsentierte auf derselben Veranstaltung ein eigenes Virtualisierungsprodukt. Ausgehend von dem freien Xen soll der „VM Server“ auf einem beliebigen x32- und x64-System laufen – eine vorherige Betriebssysteminstallation sei nicht nötig. Zur Verwaltung der virtuellen Maschine liefert Oracle ein Web-Werkzeug mit. Beide Produkte sind kostenlos, der Hersteller will lediglich mit dem Support Geld verdienen. Dafür fallen mindestens 500 US-$ jährlich an. Die Software steht unter edelivery.oracle.com/oraclevm zum Herunterladen bereit.

Preview von SQL Server 2008 mit zweidimensionalen Geodaten

Seit Mitte November bietet Microsoft eine Community Technology Preview seines SQL Server 2008 online an. Damit liegt das Produkt hinter dem Zeitplan zurück: Bei der Launch-Konferenz Ende Februar dürfte bestenfalls eine Beta-Version verfügbar sein.

Auf der Teched Developers Conference kündigte Microsoft an, dass die neue Version der Datenbank vektorielle Geodaten verwalten könne. Dabei unterscheide sie zwischen den Datentypen *Geometry* für rechtwinklige Koordinatensysteme (Gauss-Krüger, UTM et cetera) und *Geography* für geografische Koordinaten (Längen- und Breitenangaben). Eine Umrechnung zwischen beiden Systemen unter Berücksichtigung der jeweiligen Projektion soll SQL Server 2008 nicht bieten, außerdem beschränkt er sich auf zweidimensionale Koordinaten. Die üblichen Funktionen für diese Daten (Entfernung, „ist enthalten“, „schneidet“) bringt er für beide Koordinatentypen mit. Alle Versionen der Datenbank sollen Geodaten verarbeiten, einschließlich der kostenlosen Express-Variante. Mit geografischen Rasterdaten wird SQL Server 2008 jedoch nicht umgehen können, sodass das Überlagern von Luft- oder Satellitenbildern mit vektoriellen Informationen nur in der Anwendung zu bewerkstelligen ist.

Buffer-Overflow in Oracle 10gR2

Informationen von iDefense (labs.idefense.com) zufolge können authentifizierte Anwender auf Oracles Datenbankserver 10g Release 2 beliebigen Code mit den Rechten der Datenbank ausführen. Dies liege an einem Buffer-Overflow in der Prozedur *XDB.XDB_PITRIG_PKG.PITRIG_DROPMETADATA*, den präparierte *NAME*- und *OWNER*-Parameter auslösten. Aus ihnen erzeugt die Prozedur eine SQL-Anfrage, prüft deren Länge jedoch vor der Übergabe an den Server nicht.

Spezielle Privilegien seien für die Ausnutzung der Lücke nicht erforderlich. Einen Workaround gebe es bislang nicht. iDefense verwies auf einen Critical Patch Update, den der Hersteller bereitstellen müsse. Dem widersprach der Oracle-Sicherheitsspezialist Alexander Kornbrust: Die Patches vom April 2007 und später stellen seiner Meinung nach bereits eine Korrektur des Fehlers bereit. Korngold berichtete außerdem von einem öffentlichen Exploit für 10.2.01/10.2.02, mit dem sich die Datenbank zum Absturz bringen lasse. Dafür sei lediglich das *CREATE SESSION*-Privileg erforderlich.

Zugriff auf DV-gestützte Buchhaltung

Finanzbehörden haben bei Außenprüfungen von steuerpflichtigen Unternehmen ein umfangreiches Einsichtsrecht in steuerrelevante Unterlagen und Informationen. Dies umfasst unter anderem die Buchhaltung eines Unternehmens, auch wenn diese mithilfe von DV-Systemen geführt wird. Jetzt hat der Bundesfinanzhof – das in steuerrechtlichen Fragen höchste deutsche Gericht – entschieden, dass die Steuerbehörden auch auf solche elektronischen Daten zugreifen dürfen, die ihnen in Papierform vorliegen. Ein Unternehmen hatte sich geweigert, die DV-Systeme den Kontrolleuren zugänglich zu machen und stattdessen Ein- und Ausgangsrechnungen als Ausdruck auf Papier zur Verfügung gestellt. Steuerpflichtige müssen den Behörden alle elektronischen Daten der Finanzbuchhaltung zur Verfügung stellen, urteilten die Richter. Ein Steuerpflichtiger könne nicht entscheiden, ob er Daten elektronisch oder auf Papier vorlegt.

In eine ähnliche Richtung geht eine weitere Entscheidung des Bundesfinanzhofs. Ein steuerpflichtiges Unternehmen sperrte den Prüfern den Zugriff auf bestimmte Einzelkonten, die nach seiner Meinung ohnehin nur zu einer niedrigeren Steuer geführt hätten. Auch hier erteilten die Richter dem Unternehmen eine Abfuhr mit dem Hinweis, dass das Zugriffsrecht der Steuerbehörden nicht eingeschränkt werden darf – gleich aus welchen Gründen. *Tobias Haar*

Streit um Urheberabgabe für Drucker

Hersteller von Geräten, die urheberrechtlich geschützte Werke durch „Ablichtung eines Werkstücks oder in einem Verfahren vergleichbarer Wirkung“ vervielfältigen können, müssen dafür eine Abgabe zahlen. Die sogenannten Verwertungsgesellschaften treiben die Gebühren ein und verteilen sie nach komplizierten Schlüsseln an Autoren und Künstler. So steht es im Urheberrechtsgesetz. Seit Langem ist allerdings umstritten, ob auch Drucker eine Ablichtung eines Werkes ermöglichen oder nicht.

Die Gerätehersteller sind der Meinung, dass die Ablichtung schon im Rechner erfolgt und nicht erst auf dem Drucker. Aus ihrer Sicht findet die Vorschrift über Gerätevergütungen auf Drucker deswegen keine Anwendung und es ist keine Abgabe zu zahlen. Dieser Auffassung hat sich jetzt erneut das Oberlandesgericht Düsseldorf (Urteil vom 13. November 2007, I-20 U 186/06) angeschlossen.

Geklagt hatte Canon gegen die Verwertungsgesellschaft Wort. In anderen Verfahren, etwa dem von Hewlett-Packard gegen die VG Wort, hatten die Verwertungsgesellschaften Recht bekommen. Jetzt muss der Bundesgerichtshof abschließend entscheiden und damit den Streit ein für alle Mal beenden. Wenn nicht danach der Gesetzgeber noch eine Änderung beschließt, die wiederum zum Gegenstand von Gerichtsverfahren wird. *Tobias Haar*

Vorratsdatenspeicherung ist beschlossen

Mit dem Beschluss des deutschen Bundestages, die EU-Vorgaben zur Vorratsspeicherung von Telefon- und Internetdaten in deutsches Recht umzusetzen, hat die Diskussion um die ständig verschärften Sicherheitsgesetze in Deutschland einen neuen Höhepunkt erreicht. Trotz zahlreicher Proteste hat der Bundesrat am 30. November die umstrittene Gesetzesänderung genehmigt. Ab ihrem Inkrafttreten am 1. Januar 2008 müssen Provider sämtliche Telefondaten für einen Zeitraum von sechs Monaten speichern. Für Internetdaten gilt eine Übergangsfrist, sodass sie erst ab 1. Januar 2009 gespeichert werden müssen. Brisant ist in den Augen vieler Kritiker, dass es sich um eine verdachtsunabhängige Speicherung handelt. Die Daten werden auf Vorrat vorgehalten, für den Fall, dass man sie einmal benötigen sollte, beispielsweise in strafrechtlichen Ermittlungsverfahren.

Die Verpflichtung trifft Telekommunkationsanbieter und Internetprovider gleichermaßen. Dazu zählen auch Privatpersonen und alle anderen Unternehmen, wenn sie etwa Internetzugänge anbieten – also alle, die unabhängigen Dritten eine Kommunikation mit technischen Hilfsmitteln ermöglichen. Sie umfasst die Speicherung sämtlicher sogenannter „Verkehrsdaten“. Hierbei handelt es sich um Daten zu Telefonverbindungen – etwa im gleichen Umfang wie auf Einzelverbindungsnachweisen –, Daten zum Verbindungsaufbau mit dem Internet einschließlich der IP-Adressen sowie Daten zu E-Mail-Verkehr, SMS und Telefaxen. Daneben sollen auch Standortdaten und Informationen zur Geräte-Identifikation wie die IMEI-Nummer bei Mobilfunkgeräten und Funkzellenangaben erfasst und zugeordnet werden. Die Inhalte der Kommunikation können Provider auf der Grundlage dieses Gesetzes jedoch nicht speichern, da es sich bei ihnen nicht um Verkehrsdaten handelt. Vielmehr soll es den Behörden möglich sein, die gesamte elektronische Kommunikation einer Person zu analysieren.

Flatrate schützt vor Speicherung nicht

Die Gesetzesänderung setzt eine entsprechende Richtlinie über die Vorratsdatenspeicherung der Europäischen Union aus dem Jahr 2006 um, der damals außer Irland und der Slowakei alle EU-Mitgliedsstaaten zugestimmt hatten. Mit dem Gesetz ist nun auch wieder eine systematische Speicherung der Verkehrsdaten bei Flatrate-Nutzern vorgeschrieben, die zuvor mehrere Gerichte untersagt hatten. Insbesondere T-Online war mehrfach verklagt worden, weil der Konzern diese Daten gespeichert hatte, obwohl sie zu Abrechnungszwecken nicht erforderlich sind.

Sicher ist, dass sich die deutschen Gerichte bis hin zum Bundesverfassungsgericht mit den Neuregelungen auseinandersetzen müssen. Denn viele Gegner stellen die Verfassungsmäßigkeit der Bestimmungen infrage und haben bereits Klagen angekündigt. Die neuen Regelungen könnten gegen das im Grundgesetz geschützte Persönlichkeitsrecht, das Fernmeldegeheimnis, die Pressefreiheit sowie den Verhältnismäßigkeitsgrundsatz verstoßen.

Während einzelne von einem „tiefschwarzen Tag für die Bürgerrechte in Deutschland“ sprechen und bedauern, dass sämtliche Bürger unter einen Generalverdacht gestellt werden, verteidigt Bundesjustizministerin Zypries das Gesetzespaket. Nach ihrer Auffassung befindet sich Deutschland nicht auf dem Weg zu einem Überwachungsstaat. Außerdem würden nur Daten gespeichert, die ohnehin anfallen. *Tobias Haar*

Herausgabe von SMS-Spammer-Daten

Der Bundesgerichtshof (Urteil vom 19. Juli 2007, Az. I ZR 191/04) hat einen Mobilfunkanbieter dazu verurteilt, Namen und Anschrift eines SMS-Spammers herauszugeben. Das Besondere an dem Urteil ist, dass ein privater Verbraucher den Anbieter verklagt hatte. Verklagt wurde T-Mobile, weil der Verbraucher anhand der Absender-Rufnummer der Spam-Nachricht erkennen konnte, dass es sich um einen Nutzer mit einer Rufnummer aus den Nummernblöcken von T-Mobile handelte. Streitig war an diesem Fall, ob eine Vorschrift im sogenannten Unterlassungsklagengesetz auch für Empfänger ungewollter SMS gilt. Nach dieser Vorschrift besteht ein Auskunftsanspruch auf Namen und Adresse des Spammers ebenso für Verbraucher, wie die Richter jetzt klarstellten. T-Mobile wollte einen Auskunftsanspruch nur gegenüber Verbänden akzeptieren. Hintergrund ist wohl, dass Mobilfunkanbieter einen hohen Verwaltungsaufwand befürchten, wenn sich vermehrt Einzelpersonen bei ihnen wegen entsprechender Auskünfte melden. Auch die beiden Vorinstanzen hatten der Klage bereits stattgegeben, die jetzt höchstrichterlich entschieden wurde. *Tobias Haar*

Einspruch möglich bei Einführung neuer TLDs

Bis zu viertausend neue Adresszonen im Netz erwarten Domain-Fachleute als Folge der anstehenden Öffnung des Domain Name System (DNS) durch die Internet Corporation for Assigned Names and Numbers. Das schätzte deren Präsident und CEO Paul Twomey anlässlich des IGF in Brasilien (siehe S. 18). Auf der Basis von Vorschlägen des zuständigen ICANN-Gremiums, der Generic Name Supporting Organisation (GNSO), arbeitet ICANNs hauptamtliches Büro unter Twomey bereits an der Umsetzung. Als besonders heikel erweist sich die Gestaltung von Einspruchsmöglichkeiten gegen neue TLD-Begriffe.

Die GNSO hatte nicht nur Einsprüche aufgrund geografischer, Namens- und Markenrechte aufgelistet, sondern auch Einwände auf der Basis von Moral und öffentlicher Ordnung. Der frisch gewählte Chef des ICANN-Direktoriums, Peter Dengate Thrush, selbst Anwalt für Urheber- und Markenrechte, stellt klar, dass sich die GNSO mit der Formulierung an internationale Regeln angelehnt habe. Bei der Umsetzung gelte es zu beachten, dass keine böswilligen oder beliebigen Einsprüche möglich werden und den Prozess der Einführung einer neuen Adresszone blockieren. Als Anwalt werde er dazu beitragen, dass eine klare Regel zur Zulässigkeit von Einsprüchen gezogen werde.

Vertreter aus dem ICANN-Vertretergremium der nichtkommerziellen Nutzer haben davor gewarnt, dass strenge Regeln dazu führen könnten, dass am Ende nur ein kleiner Teil von neuen Domain-Endungen als unbedenklich das Verfahren passieren könnte. ICANN muss darüber hinaus Anfragen von TLD-Bewerbern aus früheren Verfahren klären. So hatte Chris Ambler, Chef von Image Online Design, den scheidenden Vint Cerf an seine Bewerbung um die TLD .web erinnert. Auch weitere ehemalige Bewerber aus der ersten Bewerberrunde für neue Domains aus dem Jahr 2000 haben ihr Interesse angemeldet. Sie haben damals immerhin 50 000 Dollar Bewerbungsgebühr bezahlt. Dieses Mal könnte es noch teurer werden. *Monika Ermert*

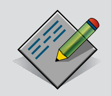
iX-Umfrage: J(2)EE klar vor .Net

Über 600 Teilnehmer konnte die – nicht repräsentative – Online-Umfrage verzeichnen, die parallel zu *iX* 12/07 auf www.ix.de lief. Es ging um den Einsatz einer Programmier- und Laufzeit-Umgebung, und zwar vor allem um die Alternative J(2)EE oder .Net. Immerhin 43 % der Antwortenden setzen auf das Java-Framework, die Alternative aus Redmond kommt nur auf ein knappes Fünftel. Bei rund jedem Siebten wird beides eingesetzt, und 24 % kommen ohne eines der beiden Frameworks aus. Die neue Umfrage dreht sich um das Ausmaß der E-Mail-Nutzung.

Wird in Ihrer Firma als Programmier- und Laufzeitumgebung Java Enterprise Edition, Microsoft .Net, beides oder keines von beiden eingesetzt?

J2EE	43 %
.Net	19 %
Beides	14 %
Keines von beiden	24 %

KURZ NOTIERT



Auditwerkzeug: Netze auf Sicherheitslücken und Malwarebefall zu überprüfen ist die Aufgabe von Pandas „Malware Radar“, der in neuer Version vorliegt.

Datensicherheit: Speziell für Behörden und Unternehmen eignet sich die neue Version 5.2 der Datensicherheitssuite SafeGuard von Utimaco mit Modul zum verschlüsselten Datenaustausch mit Kunden.

Linux-Virenschutz: Gleich drei neue Linux-Lösungen für Unternehmen bringt AV-Hersteller Eset (www.eset.de) auf den Markt, für Gateway, File- und Mailserver.

iX-Veranstaltungen

Nur noch kurze Zeit laufen die Calls for Papers für die **TeamConf 2008** und das *iX*-Cebit-Forum **Software & Systems**. Bei der erstgenannten Veranstaltung, die im April in München stattfindet, geht es, wie der Name schon andeutet, um Microsofts Visual Studio Team System. Ort und Zeit eines Cebit-Forums zu erläutern, sollte sich erübrigen. Details zu beiden Veranstaltungen sind wie immer auf der Konferenz-Website der *iX* (www.ix-konferenz.de) zu finden.

Dort kann man sich auch für die **Kerberos-Workshops** anmelden, die im Februar in Düsseldorf, München und Zürich stattfinden. Die eigentlich erst für Juni 2008 geplanten zweitägigen Competition-Roadshows zum Thema **Anforderungsmanagement** finden bereits im März statt. Einzelheiten dazu sind bei Erscheinen der vorliegenden Ausgabe ebenfalls auf der Konferenz-Site verfügbar.

www.ix-konferenz.de.

IP-Logging auch mit Flatrate

Seit Langem schwelt die rechtliche Diskussion darum, ob Internetprovider auch bei Pauschaltarifen die dynamischen IP-Adressen ihrer Kunden mitloggen dürfen. Aus datenschutzrechtlichen Gründen darf eine solche Speicherung – außer bei von staatlichen Organen angeordneten Überwachungsmaßnahmen – durch ein Unternehmen nur dann erfolgen, wenn sie zur Abrechnung des Zugangs erforderlich ist. Bei Flatrates kommt eine solche Speicherung der Nutzungsdaten also an sich nie in Betracht, da keine nutzungsabhängige Abrechnung erfolgt. Jetzt hat das Amtsgericht Bonn aber entschieden, dass eine kurzfristige Speicherung dieser Daten von bis zu sieben Tagen zur Vermeidung und Vorbeugung von technischen Störungen erlaubt ist. So sieht es eine Ausnahmevorschrift im Telekommunikationsgesetz vor. Wenn Anfang 2008 die Gesetzesänderung zur Vorratsdatenspeicherung in Kraft tritt (siehe eigenen Bericht unten), werden die Provider sogar verpflichtet sein, Nutzungsdaten sechs Monate lang zu speichern. Sie dürfen diese aber nur staatlichen Stellen bei Vorliegen bestimmter Bedingungen zur Verfügung stellen und nicht zur Störungsbeseitigung nutzen. *Tobias Haar*

Leistungsänderungen nur mit Zustimmung

Der Bundesgerichtshof hat der Praxis etlicher Internetprovider einen Riegel vorgeschoben, einseitig Inhalt und vertragliche Rahmenbedingungen eines ISP-Vertrages zu ändern. Die Richter verkannten nicht, dass sich Provider in einem dynamischen Markt bewegen, den ständige technische und inhaltliche Änderungen prägen. Deswegen ändern die Provider häufig kurzfristig die Beschreibung ihrer Leistungen. Das Recht dazu lassen sie sich von ihren Kunden per AGB einräumen. Diese unterliegen in Deutschland aber nun einmal der strengen Inhaltskontrolle durch das AGB-Recht. Trotz der hohen Veränderlichkeit auf dem ISP-Markt sei ein „allgemeiner Anpassungsvorbehalt“ des Providers unangemessen und deswegen in AGB rechtlich unwirksam, so die Richter.

Dem verklagten Provider half auch nicht, dass die Klausel vorsah, dass Leistungsänderungen dem Kunden zumutbar sein müssten. Auch eine weitere Klausel, nach der der Provider das Vertragsverhältnis durch Stillschweigen des Kunden mit einer Vorankündigung von sechs Wochen ändern darf, kassierten die Richter. Es bleibt für die Provider also nur der Weg, technische, kommerzielle und vertragliche Änderungen mit ihren Kunden über zusätzliche Verträge umzusetzen oder Vertragsänderungen wirksam zu vereinbaren, also mit ausdrücklicher Zustimmung der Kunden. *Tobias Haar*

Management vom PC-Hersteller

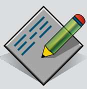
Dell nutzte Oracles Kundenmesse Openworld dafür, die Systemmanagement-Software Openmanage in der Ausgabe 5.3 vorzustellen. Sie soll insbesondere das „Klima“ in Rechenzentren positiv beeinflussen und umfasst daher Funktionen zur Überwachung des Energieverbrauchs als auch zum Monitoren virtueller Maschinen. Neu in dem kostenlosen Produkt ist unter anderem die Funktion „Inline Firmware Update“ für Aktualisierungen im laufenden Betrieb.

Integrierte Anwender- und Befehlszeilen-Schnittstellen sollen helfen, die Anzahl der Managementkonsolen zu reduzieren. Ein Skriptgenerator und die Funktion „Custom Update Policies“ beschneiden den Aufwand fürs Inbetriebnehmen und Aktualisieren von Programmen. Weiter ausgebaut hat Dell auch das Managementportfolio zum Verwalten von PC-Infrastrukturen. Nachdem man bereits im Juli durch den Kauf von Silverback in das Service-Geschäft eingestiegen war, wird das Angebot der Remote-Managementdienste nun mit der Übernahme von Everdream komplettiert. Damit steht der PC-Bauer nach Analyse der Berater von Ovum aber auch vor der Aufgabe, die agentbasierende Everdream-Lösung und die agentenlose Silverback-Software zu integrieren.

NetIQ steigt in Prozessautomation ein

Die Attachmate-Tochterfirma NetIQ steigt mit Aegus in das Segment der IT-Automatisierung ein. Die Plattform soll die bislang in der Administration manuell durchgeführten IT-Prozesse und -Routinen automatisieren und die schrittweise Einführung von „Best Practices“ nach ITIL unterstützen. Eine Integration anderer Managementlösungen erfolgt über den Enterprise Service Bus von Aegis. Eine Correlation Engine erfasst darüber Ereignismeldungen unterschiedlicher Herkunft, korreliert sie und stößt die notwendigen Prozesse gemäß den Workflow-Vorlagen automatisch an. Die Vorlagen führt eine Process Automation Engine aus. Prozessabläufe lassen sich mithilfe einer Design-Umgebung modellieren. Eine Webkonsole visualisiert die Prozessabläufe, während die Analyse-Komponente Einblicke in die Effizienz der Durchführung geben soll.

KURZ NOTIERT



Prozessfan: BMC bringt ein neues Werkzeug für die Prozessplanung auf den Markt. Das Service Management Process Model ermöglicht ein visuelles Mapping von ITIL-Praxisbeispielen mit den detaillierten Prozessen und Arbeitsanweisungen der Anwendungen fürs Business Services Management der US-Firma.

Ausgelagert: CA hat die Forschung und Entwicklung der eigenen Schutzprogramme (Anti-Virus, Anti-Spyware, Anti-Spam, Firewall etc.) an die indische HCL Technologies abgegeben. Der US-Firma bleiben Vertrieb und Marketing. Den Umsatz wollen sich beide Firmen künftig teilen.

Virtuelle Einkäufe: Quest Software übernimmt Provision Networks, einen Anbieter von Desktop-Virtualisierungslösungen. Bereits im Sommer 2007 hatte die US-Firma mit Vizioncore und Invirtus zwei auf Virtualisierung spezialisierte Unternehmen gekauft.

Gevoipt: Realtech hat ein Integrationsmodul für Ciscos Callmanager (CCM) vorgestellt. Es normalisiert CCM-Daten im theGuard Networkmanager und überführt sie in die zentrale Configuration Management Database (CMDB). Die vom CCM verwalteten VoIP-Netzwerke lassen sich so visualisieren und überwachen.

Sensoren mit RFID-Tags kombiniert

Wenn Fahrzeugteile, Flachbildschirme, Medizin oder Lebensmittel verrostet oder verdorben beim Großhändler ankommen, sind die Fehler in der Warentransportkette und der Verursacher des Schadens oft schwer zu finden. Mit einer Erweiterung von RFID-Transpondern um Sensorfunktionen soll die Frachtüberwachung in den Bereichen Pharma, Automotive und Luftfracht effektiver werden. Daran arbeiten derzeit sechs Industrieunternehmen unter der Federführung des Fraunhofer-Instituts für Physikalische Messtechnik IPM im Verbundprojekt TRACK („Traceability: Rückverfolgbarkeit durch Autonome Mikrosysteme zum kontinuierlichen Check von Konsumgütern“; www.track-projekt.de).

Ziel ist eine RFID-Plattform mit standardisierten Schnittstellen, in die Unternehmen je nach Bedarf verschiedene, eigens für diese Anwendung entwickelte Sensoren integrieren können. So überwachen Feuchtsensoren rostempfindliche Autoteile und Temperatursensoren prüfen die Kühlung von Medikamenten. Licht- und Beschleunigungssensoren wachen darüber, ob Luftfrachtcontainer während des Transports geöffnet oder stark erschüttert wurden. Während des Warentransports nehmen die Sensoren kontinuierlich Messwerte auf und speichern sie samt Zeitmarken im RFID-Chip, den der Empfänger der Ware per Funk berührungslos auslesen kann.

Im Falle einer Reklamation lässt sich die Ursache des Schadens leicht ermitteln. Eine besondere Herausforderung liegt nach Angabe der Forscher in der Entwicklung energiesparender und gleichzeitig kostengünstiger Sensoren. Erste Prototypen sollen im kommenden Jahr in der Praxis erprobt werden. *Barbara Lange*

SSH für sichere FTP-Übertragungen

SSH Communications Security (www.ssh.com) hat Tectia 6.0 um ein neues Produktmitglied namens ConnectSecure erweitert. Es soll FTP-Dateitransfers und Datenübertragungen absichern, ohne dass Modifikationen in Anwendungen, Skripts oder Infrastruktur erforderlich sind. Dafür bietet die Software automatische FTP- zu SFTP-Transformation oder transparentes FTP- und TCP-Tunneling. ConnectSecure läuft auf Unix-, Linux- oder Windows-Plattformen und soll kompatibel sein zu anderen SSH-Serverlösungen wie OpenSSH oder solchen von Drittanbietern. Version 6.0 umfasst zudem einen Windows-Client, der auch transparentes TCP-Tunneling sowie automatisches Tunneling durchführt. Neu ist außerdem die Unterstützung von Fast Streaming Crypticore Ciphers auf Unix- und Linux-Plattformen mit Intel-Architektur. Zur Tectia-Produktfamilie gehören neben ConnectSecure und den Clients ein Server, der mit der neuen Version für z/OS verfügbar ist, sowie ein Manager für die zentrale Verwaltung. Die Produkte sind ab Januar 2008 erhältlich, der Server 6.0 für z/OS ab März. *Susanne Franke*

KURZ NOTIERT

Logdatenverwaltung: syslog-ng von Balabit (www.balabit.com), zentrales Werkzeug zur Sammlung und Verwaltung von Logdaten zur Systemüberwachung, unterstützt nun auch IBMs System-i-Plattformen (vormals AS/400). Überdies soll das Werkzeug das Einhalten von Datenschutzrichtlinien und -standards unterstützen.

Leoparden-Scanner: AV-Spezialist McAfee (www.mcafee.com) bringt mit VirusScan für Mac v8.6 einen Virenschutz für das neue Apple-Betriebssystem heraus. Das Scan-Programm lässt sich über McAfees Managementkonsole zentral steuern und verwalten.

TOP 20 der Sicherheitsrisiken

Ende November veröffentlichte das SANS-Institut wie in jedem Jahr die 20 größten IT-Sicherheitsrisiken nebst Entwicklungen im Vergleich zum Vorjahr. Signifikant zugenommen hat die Bedrohung durch clientseitige Schwachstellen, an erster Stelle steht hier der Webbrowser. Ungepatchte oder veraltete Versionen, aber ebenso Plugins von Drittanbietern, die Active Scripting zulassen, bieten Cyberkriminellen Angriffsflächen – das gilt für den am meisten verbreiteten Browser IE ebenso wie für Platz 2, Mozillas Firefox. Auch haben die Risiken bei Office-Anwendungen spürbar zugenommen, während die Betriebssysteme deutlich weniger anfällig für Internetwürmer als noch in den Jahren davor sind. Unangefochten auf Platz 1 bei den Serveranwendungen sind wie schon im vergangenen Jahr Web Applications. Die weiteren Gefahren, Trends und praktische Hinweise zur Risikovermeidung finden sich auf der SANS-Website (www.sans.org).

Mobilfunkgeräte werden zu RFID-Lesern

Samsung Electronics hat einen RFID-Lese-Chip für mobile Geräte entwickelt. Als Haupt-Einsatzgebiet sieht das Unternehmen Kartenlesegeräte, die sich in Mobiltelefone einstecken lassen. Damit wird das Handy zum RFID-Lesegerät. Endverbraucher können damit zum Beispiel Produkt- und Service-Informationen von Plakaten, Eintrittskarten oder Kleidungsstücken, die mit einem RFID-Transponder ausgerüstet sind, auf ihr Handy laden. Der 6,5mm × 6,5 mm große Chip funkt im UHF-Band bei 900 MHz, enthält eine analoge Eingangsstufe, ein Basisband-Modem, einen Prozessor sowie einen Speicher. Er ist ab sofort verfügbar. Nach Untersuchungen der Marktforscher der Research on Asia (RoA) Group soll der weltweite Bedarf an RFID-Chips für Mobilgeräte von derzeit 26,9 Milliarden US-Dollar auf 701 Milliarden im Jahr 2010 steigen – das entspricht einem durchschnittlichen Wachstum von 196 Prozent pro Jahr. *Barbara Lange*

Webanwendungs-Firewall lernt mit

Der Schweizer Hersteller Visonys (www.visonys.ch) hat seine Webanwendungs-Firewall Airlock in Version 4.1 um ein Monitoring- und Reporting-System erweitert. Eine zentrale Log-Engine bearbeitet die verschiedenen Protokollinformationen. Über eine Suchmaske kann der Mechanismus die Log-Meldungen nach Herkunft, Kategorie und Schweregrad filtern, die dann über eine Weboberfläche angezeigt werden. Bei vordefinierten Events gibt das System eine Alarmmeldung aus. Ein grafisches Reporting-Modul soll zudem die Auswertung von Messdaten wie Antwortzeiten, Status-Codes oder Filterergebnissen ermöglichen. Eine weitere Änderung betrifft die Antiviren-Prüfprozesse: Eingehende Anfragen zur Erkennung von Viren und Malware sendet die Firewall über eine ICAP-Verbindung (Internet Content Adaption Protocol) an zuvor festgelegte Antiviren-Gateways, die sie dort überprüfen. Die Web Application Firewall verhält sich dabei wie ein ICAP-Client. Die Software ist applikationsunabhängig und kontrolliert den Zugang zu allen Webanwendungen durch plattformübergreifendes Single-Sign-on. Schließlich können Anwender in der Version 4.1 von Airlock wählen, ob die Firewall statisch voreingestellte URLs erkennen oder ob sie dynamisch dazulernen soll. *Susanne Franke*

Praxiserfahrung gefragt

Deutschlands IT-Branche bietet Hochschulabsolventen gute Einstiegschancen – aber nur denen mit Berufserfahrung. Für fast 80 % aller ausgeschriebenen Einstiegspositionen fordern IT-Unternehmen zusätzlich Berufserfahrung von den Bewerbern. Praktika alleine genügen nur für gut jede zehnte offene Stelle. Für die Mehrzahl aller freien Einstiegsjobs erwarten Arbeitgeber dagegen tiefer gehende Erfahrungen aus dem Berufsalltag. Knapp 30 % fordern sogar erste Erfahrungen im Management oder beim Umsetzen von Projekten. Zu diesen Ergebnissen gelangte die aktuelle Berufsstudie IT-Jobscout 2007, für die der IT-Dienstleister PPI AG insgesamt 624 Stellenanzeigen der 100 größten deutschen IT-Unternehmen auswertete.

Berufsbegleitender Master-Studiengang

Die Zentralstelle für Fernstudien an Fachhochschulen (ZFH), die für die Fachhochschulen der drei Bundesländer Hessen, Rheinland-Pfalz und Saarland Fernstudien entwickelt, stellte einen berufsbegleitenden Master-Studiengang Elektrotechnik vor. Für das dreijährige Fernstudium sind derzeit die beiden Vertiefungsrichtungen Automatisierung und Mikroelektronik vorgesehen. Die Studierenden erhalten nach erfolgreichem Studium den international anerkannten akademischen Abschluss eines Master of Science (M. Sc.). Mit dem Master steht ihnen sowohl der Weg zur Promotion als auch zum höheren Dienst bei öffentlichen Arbeitgebern offen. Interessenten für das Fernstudium müssen ein einschlägiges Hochschuldiplom beziehungsweise Bachelor oder anerkannten äquivalenten Abschluss nachweisen. Als einschlägig gelten Elektrotechnik, Mechatronik sowie Technische Informatik. Zusätzlich wird eine mindestens einjährige Berufspraxis nach Abschluss des Erststudiums verlangt. Zulassungsanträge sind für das Sommersemester bis zum 15. Januar zu stellen. Unterlagen erhält man unter www.zfh.de/informationen/anmeldung.

Ohne Abitur zum Studium

Bewerbern ohne ausreichende Hochschulreife räumt die Fernuni Hagen die Möglichkeit ein, den Zugang zu einem Studiengang mittels der Zugangsprüfung zu erwerben. Interessenten, die sich in der beruflichen Bildung qualifiziert haben, wird dies über eine Zugangsprüfung eingeräumt. Voraussetzungen sind ein Mindestalter von 22 Jahren, eine abgeschlossene Berufsausbildung sowie eine mindestens dreijährige Berufstätigkeit. Vorab müssen allerdings im sogenannten Akademiestudium Leistungen für den angestrebten Studiengang, etwa Informatik oder Wirtschaftsinformatik, erbracht werden. Die nächste Bewerbungsfrist für die Teilnahme an der Zulassung als Akademiestudierende/r endet am 15. Januar. Weitere Infos unter www.fernuni-hagen.de/studium/vor/bewerbung/zugangspruef.shtml.

KURZ NOTIERT

Löblich: IBM erhöht die Anzahl der Ausbildungsplätze an der Berufsakademie (BA) ab 2008 um 60 auf 240. An der BA werden neben dem Studiengang Wirtschaftsinformatik auch Angewandte Informatik, Dienstleistungsmanagement und International Business angeboten.

Neue Onlinebörse: Unter www.critex.com wurde eine neue Job- und Projektbörse gestartet, die osteuropäische IT-Experten und hiesige mittelständische Firmen zusammenbringen soll.

Studie zum TK-Markt 2007

Leicht rückläufig

Achim Born

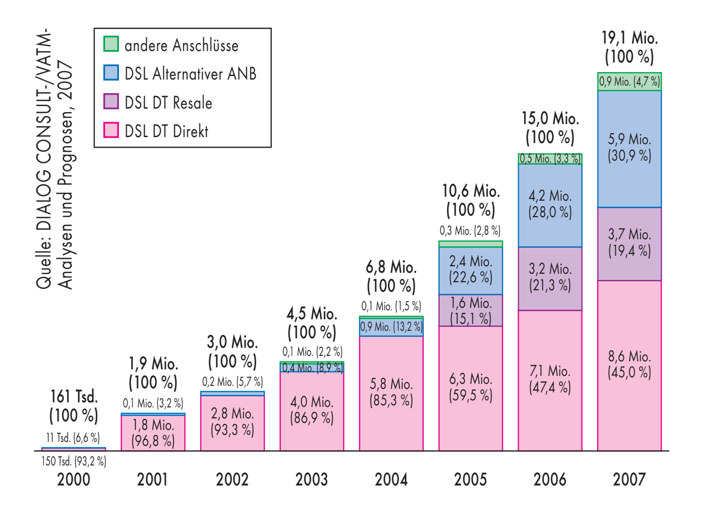
Die Telekommunikationsanbieter müssen sich auf rückläufige Geschäfte einstellen. Mengenwachstum und DSL-Boom können den Preisverfall nicht ausgleichen.

Die Umsätze mit TK-Diensten in Deutschland werden sich in diesem Jahr auf 63,4 Mrd. Euro belaufen. Das entspricht im Vergleich zum Vorjahr einem Rückgang um 2,8 %, obgleich die schiere Zahl der Verbindungsminuten in Festnetz und Mobilfunk weiter steigt. Dies geht aus der von VATM und Dialog Consult GmbH vorgestellten zehnten gemeinsamen TK-Marktstudie hervor. Von den Gesamtumsätzen entfallen 37 Mrd. Euro auf den Festnetz- und 26,4 Mrd. Euro auf den Mobilfunkmarkt. Bei den Komplettanschlüssen konnten VoIP-Anschlüsse den Anteil an den Verbindungsminuten verdoppeln. Damit entfallen in diesem Jahr täglich 55 Mio. Minuten auf IP-Netze. Weitere 11 Mio. Verbindungsminuten täglich laufen über TV-Breitbandkabel.

Nach Einschätzung der Studienautoren steigt die Zahl der direkt geschalteten Breitbandanschlüsse bis zum Jahresende um 27 % auf 19,1 Millionen. Dabei bleibt Deutschland DSL-Land. Breitbandanschlüsse über alternative Anschlussarten wie Kabelmodem, Powerline oder Satellit konnten ihren Anteil gerade einmal von 3,3 auf 4,7 % ausbauen. Ihre Anzahl bleibt damit unter der magischen Millionen-Grenze. Zwei von drei Personen mit Breitbandanschluss nutzen Downstreams zwischen 2 und 6 MBit/s. Bei jedem Zwölften geht es noch schneller. Jeder Vierte begnügt sich jedoch mit maximal 2 MBit pro Sekunde. Das jährliche Breitband-Verkehrsvolumen wird voraussichtlich 2007 weiter auf 1,023 Mrd. Gigabyte steigen. Das durchschnittliche Datenvolumen pro Nutzer ist jedoch weiterhin rückläufig, da der Anteil der „heavy user“ im Verhältnis zur steigenden Gesamtnutzerzahl sinkt. Jeder Nutzer generiert zurzeit laut Untersuchung pro Monat ein durchschnittliches Datenvolumen von 4,5 GB.

Unangefochten die Nummer eins im um gut 3,6 % rückläufigen Festnetzmarkt ist die Deutsche Telekom mit einem Umsatzanteil von 62,5 %. Allerdings konnten die Wettbewerber ihre Einnahmen im Festnetz um 0,6 Mrd. Euro entsprechend 4,5 % auf knapp 14 Mrd. Euro steigern. Dieses Wachstum erzielen sie primär im Bereich der Breitbandanschlüsse, der weiterhin boomt. Da die Konkurrenten aber an die Deutsche Telekom sogenannte Vorleistungsentgelte zahlen müssen, tragen sie nach wie vor erheblich zu den Umsätzen des ehemaligen Staatsunternehmens bei.

So sind bei einem Komplettanschluss von jedem Umsatz-Euro 52 Cent, bei einem Resale-Anschluss der Telekom von jedem Euro sogar 91 Cent an die Telekom weiterzureichen. Unter Berücksichtigung dieser „Wertschöpfungskette“ ist der Ex-Monopolist auf dem boomenden DSL-Markt ebenfalls der große Gewinner. Denn der Zuwachs in absoluten Zahlen liegt mit zwei Mio. inklusive der Reseller-Anschlüsse deutlich höher als der der alternativen Festnetzanbieter, die 2007 gemeinsam auf gerade einmal 1,7 Mio. neue Anschlüsse kommen.



Kein Morgen für Tomorrownow?

Mehrere Führungskräfte von Tomorrownow, darunter auch der Chef Andrew Nelson, müssen die SAP-Tochter verlassen. Des Weiteren prüft das SAP-Management nach eigenen Angaben verschiedene Zukunftsszenarien; ein Verkauf wird explizit nicht ausgeschlossen. 2005 hatte der Walldorfer Softwareriese die US-Firma Tomorrownow erworben, um mithilfe kostengünstiger Wartungsangebote für Peoplesoft, J. D. Edwards, Siebel et cetera im Anwender-Terrain des Konkurrenten Oracle zu wildern.

Völlig unvorbereitet traf die SAPler jedoch dann der Vorwurf der Industriespionage, den Oracle öffentlichkeitswirksam erhob und der letztlich im Frühjahr 2007 in einer Anzeige mündete. Nachdem zu Beginn das SAP-Management, allen voran SAP-Chef Henning Kagermann, noch sehr kämpferisch auftrat und den Vorwurf brüsk abwies, musste es schlussendlich doch kleinlaut ein Fehlverhalten bei der Tochterfirma einräumen. Konkret wurden „nicht angemessene Downloads“ von Oracle-Servern durch Tomorrownow eingestanden. Mit dem jetzigen Schritt zieht man wohl die späte Konsequenz aus dem damaligen Fehlverhalten. Auf den Prozessverlauf im Rechtsstreit hat die neue SAP-Politik jedoch erst einmal keine direkte Auswirkung. Hier steht noch das Ende der Beweisermittlung an.

United Internet bringt sich in Position

Die United Internet AG (u. a. 1 & 1 und GMX) baut sukzessive ihre Stellung im hiesigen TK-Markt aus. Ende November sicherte sich das Montabaur Unternehmen für rund 127 Mio. Euro einen Anteil von 20,05 % an Versatel inklusive Aufstockungsoption. Attraktiv macht den seit Längerem als Übernahmekandidat gehandelten Konkurrenten die eigene Netzinfrastruktur. Unter anderem beabsichtigt die Düsseldorfer Firma den Aufbau eines VDSL-Netzes in Konkurrenz zur Deutschen Telekom.

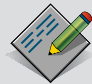
Groß geworden ist das Unternehmen allerdings ohne eigene Infrastruktur, als Reseller von DSL-Anschlüssen der Deutschen Telekom. Seit Juli bietet man auch Komplettpakete inklusive Telefonanschluss an und greift hierfür auf das Netz von Telefonica und QSC zurück.

Auf jedem Fall besitzt United Internet nun alle Optionen, an der Konsolidierung des deutschen Internet-Provider-Marktes zu partizipieren. In diesem Kontext ist auch der kurz zuvor erfolgte Einstieg bei Drillisch zu beurteilen. Mit dem Mobilfunker beabsichtigt man bereits seit geraumer Zeit, die Hamburger Freenet zu übernehmen, auch wenn United Internet offiziell die Verhandlungen für gescheitert erklärte. Über eine gemeinsame Holding GmbH hält man mit Drillisch nun bereits mehr als 10 % der Anteile an dem Hamburger Unternehmen.

Neuausrichtung bei Onvista

Die Onvista Group will sich künftig auf Betrieb und Vermarktung des gleichnamigen Finanzportals inklusive der Börsen-Community Tradingbird konzentrieren. Alle anderen Geschäftsbereiche – das Gesundheitsportal Onmeda, der Online-Werbevermarkter Ad2Net, der Performance-Marketing-Dienstleister Ligatus und das Adress-Brokerage-Portal Namendo – stehen zum Verkauf an. Hintergrund des Strategiewechsels sind die neuen Mehrheitsverhältnisse in der Unternehmensgruppe. Die französische Boursorama SA, die inzwischen über 80 % der Anteile an der AG hält, strebt eine vollständige Übernahme an und hat den übrigen Aktionären ein Übernahmeangebot unterbreitet. Das französische Unternehmen plant nach seinen Online-Broker Fimatex und das werbefinanzierte Finanzportal Onvista mittelfristig eng miteinander zu verzahnen.

KURZ NOTIERT



Eingekauft: Intuit, Hersteller der gleichnamigen Finanzmanagementsoftware, erwirbt Homestead Technologies. Der Kaufpreis für die US-Firma, die unter anderem als Hoster von Websites für Unternehmen aktiv ist, soll dem Vernehmen nach rund 170 Mio. Dollar betragen.

Ausgebaut: Die Cundus AG hat AIS Consulting übernommen. Durch den Kauf erweitert das Duisburger Unternehmen die Kompetenzen im Bereich der betriebswirtschaftlichen Planung und Konsolidierung auf Basis von SAP- und Microsoft-Produkten. AIS Consulting verfügt unter anderem über langjährige Erfahrungen mit SAP BPC Business Planning and Consolidation.

Mainframe-Fan: Beta Systems übernimmt sämtliche Geschäftsanteile der SI Software Innovation GmbH. Durch die Akquisition wollen die Berliner ihr Mainframe-Geschäft, insbesondere im Bereich Output-Management und Archivierung, ausbauen.

Zufrieden: Die britische Sage Group steigerte im Geschäftsjahr 2007 (endete im September) den Umsatz hierzulande um 36 % auf 79,7 Mio. Euro. Der operative Gewinn (EBITA) stieg um rund 31 % auf 16,5 Mio. Euro. Neben dem organischen Wachstum, nach Eigenangabe 10 % geht der Zuwachs auf die im Juni 2006 erfolgte Übernahme des Konkurrenten Bäurer zurück.

BI-Spezialisten in Not: Nach Business Objects (SAP) und Hyperion (Oracle) verliert nun Cognos die Eigenständigkeit. Rund fünf Mrd. Dollar in bar will es sich IBM kosten lassen, sich den kanadischen Software-Anbieter einzuverleiben. Im ersten Quartal des neuen Jahres soll der Kauf abgeschlossen sein.

Anstieg: Die deutsche Industrie erzielte eine neue Höchstmarke beim Export von Software und ITK-Diensten. Im ersten Halbjahr 2007 stiegen laut Bitkom die Ausfuhren um 3 % auf 5,9 Mrd. Euro. Parallel zum Anstieg der Exporte haben die Importe von Software und ITK-Diensten um 5,3 % auf 6,3 Mrd. Euro zugelegt. Der größte Anteil kommt mit fast 20 % aus den USA. Das stärkste Wachstum verzeichnen Einfuhren aus Indien, allerdings auf einer niedrigen Basis.

Begehr: HP hat Appetit auf ein größeres Stück am Softwaremarkt. Es sind nach Insider-Informationen Übernahmen im Umfang von mehreren Mrd. Dollar geplant. Bereits in den vergangenen anderthalb Jahren hatte HP unter anderem die Softwarefirmen Bristol, Mercury, Opsware, Peregrine und SPI geschluckt.

Gartners Top-10-Techniken für 2008

1. Green IT	6. Mashup & Composite Apps
2. Unified Communications	7. Webplattform & WOA
3. Geschäftsprozessmodellierung	8. Computing Fabric
4. Metadaten-Management	9. Real World Web (Einsatz von Echtzeitdaten)
5. Virtualisierung 2.0	10. Social Software

Gartners Top 10 der IT-Trends 2008

Wer auch im neuen Jahr an den CIO-Stammtischen mitreden will, kommt um Gartners Top-10-IT-Trends nicht herum.

Wieder einmal haben die Gartner-Analysten ihrer alljährlichen Fleißarbeit gefrönt und die zehn „strategischen Technologien“ für IT-Manager zusammengetragen. Mit dem Adjektiv „strategisch“ adeln die US-Berater technische Entwicklungen, denen sie signifikanten Einfluss auf die Unternehms-IT attestieren. Die veröffentlichte Liste will man allerdings eher als Überlegungs- denn als Handlungsempfehlung verstanden wissen. Die IT-Veranwortlichen sollten nach Ansicht der Berater die Technologien in ihren Planungsprozess einbeziehen und entscheiden, ob und wann die Einführung für ihre Organisation sinnvoll ist.

Wenig überraschend erkor Gartner „Green IT“ zum wichtigsten Thema. IT-Manager werden wohl aufgrund gesetzlicher Vorgaben nicht umhinkommen, sich intensiv mit der CO_2-Emission ihrer Rechenzentren und IT-Infrastrukturen zu befassen. Als Nummer 2 der zehn wichtigsten Technikthemen folgt die Vereinheitlichung der Kommunikation (Unified Communications). Die Berater umschreiben damit die breite Einführung der IP-Telefonie als Ersatz der traditionellen Festnetztelefonie und klassischen CTI-Lösungen. 20 % der Unternehmen sind nach Einschätzung von Gartner bereits umgestiegen, 80 % sammeln gegenwärtig im Rahmen von Testprojekten Eindrücke von entsprechenden Anwendungen.

Die nächsten beiden Listenplätze belegen Techniken, die sich mit der Einführung neuer prozess- und serviceorientierter Anwendungen befassen. Dies betrifft zum einen den organisatorisch-personellen Part. Hier fordert Gartner, in den Unternehmen neue Tätigkeitsfelder (Enterprise-Architekt, Prozessarchitekt oder Prozessanalyst) zu etablieren. Zum anderen raten die Analysten, die unterschiedlichen Initiativen zum Stammdatenmanagement (Meta Data Management – MDM) in den Bereichen Vertrieb, Fertigung und Kundenmanagement bis 2010 in einem umfassenden EIM-Ansatz (Enterprise Information Management) zu bündeln. Dabei gehören für sie ausdrücklich Service Registries oder Anwendungsentwicklungs-Repositories in den Einflussbereich eines solchen EIM.

Ohne Open Source

Mit Mashups und Composite Apps (Platz 6) sowie Web-Plattform und WOA (weborientierte Architektur, Platz 7) führt Gartner weitere Softwarekonzepte aus dem Bereich der Composite Apps und SOA als bedeutende technische Trends an. Komplettiert wird die Liste von den Punkten Virtualisierung, Weiterentwicklung der Blade-Technik, dem sofortigen Nutzen von Echtzeitdaten aus dem Web und Web-2.0-Software (Social Software).

All dies war, wenn auch manchmal mit anderem Zungenschlag oder leicht unterschiedlicher Schwerpunktsetzung, bereits in der Vorjahresliste vertreten. Einzig die seinerzeitige Nummer 1 „Open Source“ fand heuer keine Aufnahme mehr.



Korrekter Umgang mit E-Mail-Adressen

Vorsorge-maßnahmen

Bert Ungerer

Berge von Spam, gefräßige Filter und juristische Unwägbarkeiten halten Internet-Anwender nicht davon ab, E-Mail für persönliche Nachrichten und als Werbeträger zu nutzen. Sie müssen jedoch Vorsicht walten lassen, sollen ihre Mails weder im Abwehr-dickicht hängen bleiben noch im Unerwünschten untergehen.

Nachrichtentechnisch betrachtet kann das Medium E-Mail (derzeitiges Signal-Rausch-Verhältnis: 1 zu 10) eigentlich schon lange nicht mehr funktionieren, aber eine Alternative scheint nicht in Sicht. Den Anbietern von allerlei Filterverfahren ist es recht: Die Mailsysteme der von Spam geplagten Empfänger sortieren mit ihrer Hilfe bereits bis zu über 90 Prozent aller E-Mails als unerwünscht aus. Selbst damit sind die Eingangsordner der Anwender selten spamfrei, enthalten aber wenigstens überwiegend Erwünschtes.

Einer der Gründe für den massiven Anstieg des Spam-Volumens in den vergangenen Jahren ist – neben den extrem niedrigen Kosten für die Versender – ausgerechnet die gestiegene Filterqualität. Je mehr Werbemüll auf dem Transportweg herausfliegt, desto mehr Ausstoß produzieren die Spam-Versender, damit sich ihre Umsatzziele erfüllen. Für einen Anwender, dessen Filtersysteme 99 Prozent der Spam-Mails entsorgen, müssen 100 Müllmails auf die Reise gehen, damit er nur eine einzige erhält.

Zum Glück sind die Empfänger dem nach wie vor ungebremst wachsenden Beschuss durch Botnetze und gekaperte Kontos bei Yahoo, Google, Hotmail und anderen Mail-Providern nicht ganz schutzlos ausgeliefert. Postmaster setzen Blacklists ein, um E-Mails aus dubiosen Quellen erst gar nicht in die eigenen Systeme gelangen zu lassen. Zentral eingerichtete Filtersysteme können massenhaft eingehende Mails erkennen und aussortieren, und bei manch einem Provider hilft es sogar, sich über Spam-Versender in dessen Netz zu beschweren.

All diese Maßnahmen haben leider einen entscheidenden Nachteil: Sie setzen erst dann an, wenn der Spam bereits unterwegs ist und wertvolle IT-Ressourcen verbraucht hat. Vorbeugung spielt daher ebenfalls eine wichtige Rolle, führt aber auch ein Schattendasein – über nicht vorhandenen Spam lassen sich keine spektakulären Statistiken führen. Einige empfehlenswerte, aber auch einige kontraproduktive Maßnahmen zur Spam-Vermeidung seien im Folgenden zusammengefasst. Sie betreffen nicht nur die Empfängerseite, denn zum Spam-Versender kann jeder mehr oder weniger unfreiwillig werden, ob Online-Marketier oder privater Internet-Anwender, der noch nicht einmal eine eigene E-Mail-Adresse hat.

Adressen im Web verschleiern

Viele „bespammte“ Adressen stammen von Webseiten, auf denen sie zum Anbahnen von E-Mail-Kontakten veröffentlicht sind oder waren. Seit das Kind in den Brunnen gefallen ist, kursieren Tipps, die Adressen etwa durch willkürliche Zusätze wie „-removethis-“ zu verschleiern oder durch eine Grafik zu ersetzen. Nicht jeder potenzielle Kunde oder Interessent kann oder will jedoch die korrekte Adresse erraten oder Text in einer Grafik erkennen und abtippen.

Ebenso wirkungsvoll, doch wesentlich bequemer für Besucher der Website,

Was tun mit Mail-Irrläufern?

Der Liebesbrief von Claudia, die Rechung an Peter oder gar die Zugangsdaten von Familie Beckers neuem DSL-Anschluss: Ein kleiner Vertipper bei der Eingabe der Mail-Adresse, und schon landet die Sendung entweder im digitalen Nirwana oder gar bei einem Empfänger, für den sie nicht bestimmt ist. Während der erste Fall durch die entstehende Fehlermeldung in der Regel unproblematisch ist, stellt sich bei einem falschen Empfänger die Frage, ob und in welchem Umfang dieser verpflichtet ist, auf den Irrläufer zu reagieren.

Zunächst einmal ist der Empfänger einer E-Mail grundsätzlich berechtigt, diese auch zu öffnen und sich Kenntnis von ihrem Inhalt zu verschaffen. Eine Verpflichtung, auf den Inhalt der E-Mail zu reagieren und die Beteiligten über die Fehlsendung zu informieren, dürfte dagegen im Normalfall nicht bestehen. Dabei spielt es auch zunächst keine Rolle, ob es sich bei dem Inhalt der Mail um einen einfachen Gruß, eine Mahnung oder eine Bestellbestätigung handelt.

Eine solche Verpflichtung könnte sich allenfalls dann ergeben, wenn der Empfänger aufgrund einer vertraglichen oder sonstigen Verbindung zur Wahrung bestimmter Fürsorgepflichten gegenüber dem Partner verpflichtet ist. Eine solche Pflicht könnte sich etwa bei einem Provider ergeben, der zufällig eine tatsächlich an einen Kunden gerichtete Mail erhält. In diesem Fall wäre er aufgrund seiner vertraglichen Pflichten zu einer Weiterleitung der Mail verpflichtet. Für private Empfänger oder sonstige Dritte, die einen Fehlläufer erhalten, gilt dies aber nicht. Hier gebietet es allenfalls die Höflichkeit, einen der Beteiligten von dem Fehler zu informieren.

An dieser Rechtslage ändern auch die häufig verbreiteten sogenannten „Disclaimer“ in E-Mails nichts. Auch wenn sie dem Empfänger Rechtsverbindlichkeit suggerieren sollen, ist ihre Wirkung nach deutschem Recht nahezu völlig unbedeutend. Eine Wirksamkeit hätten derartige „Voodoo-Klauseln“ nach hiesiger Rechtslage allenfalls dann, wenn zwischen Sender und Empfänger ein Vertrag geschlossen würde, der den Empfänger zu bestimmten Handlungen verpflichtet. Erklärungen in Disclaimern entfalten aber als einseitige Willenserklärungen in aller Regel keine Vertragswirkung und sind damit für den Empfänger bedeutungslos.

Eine Verpflichtung, auf Irrläufer zu reagieren, könnte sich allenfalls – wenn auch in unwahrscheinlichen und schwer nachweisbaren Fällen – aus dem Strafrecht ergeben. So stellt § 138 StGB die Nichtanzeige geplanter Straftaten unter Strafe. Danach wird derjenige mit einer Freiheitsstrafe von bis zu fünf Jahren bestraft, der es unterlässt, die Behörden oder die Betroffenen rechtzeitig von geplanten Straftaten zu informieren. Wie sich jedoch schon aus dem hohen Strafrahmen ergibt, gilt dies nicht uneingeschränkt für jede Verfehlung. Von dieser Vorschrift erfasst sind vielmehr nur die sogenannten „Kapitalstraftaten“ wie Mord, Totschlag, Raub oder ähnlich geartete Delikte. Eine besondere Ausnahme gibt es schließlich im Bereich der Verbreitung von Kinderpornografie. Da hier bereits der Besitz derartiger Bilder strafbar ist, sollte sich ein Empfänger von E-Mails mit solchen Inhalten immer an die Strafverfolgungsbehörden wenden, um nicht selbst in den Verdacht zu geraten, sie angefordert zu haben.

Joerg Heidrich

sind HTML-Umkodierungen der Adresse oder kleine Javascript-Programme, die die korrekte E-Mail-Adresse überhaupt erst im Webbrowser entstehen lassen. Damit lassen sich zumindest simplere Adress-Sammel-Programme (sogenannte Harvester) austricksen. Ein Hilfsmittel für solche Umkodierungen findet sich zum Beispiel unter www.antispam.de/?topic=encoder.

Leider sind Webseiten bei Weitem nicht die einzige Quelle von Adressen (siehe unten). Sie scheinen in letzter Zeit sogar an Bedeutung abzunehmen, denn im Web lauern mittlerweile zahlreiche Harvester-Fallen. Dennoch lohnt es sich, ein wenig Zeit in das Gestalten der mailto:-Verweise auf den eigenen Webseiten zu investieren. Erstaunlicherweise nutzen die meisten Webmaster sie nur zum Angeben der nackten E-Mail-Adresse. Doch wenn man schon nicht verhindern kann, dass Spammer sie erfahren, so ist es mithilfe des mailto:-Tag immerhin möglich, zusätzliche Informationen einzubauen, die Spammer ignorieren und die der eigene Mail-Filter daher als Positivmerkmal lernen kann.

Ein praktisches Beispiel dazu findet sich auf der Informationsseite zur *iX*-Blacklist www.ix.de/nixspam/dnsbl/. Ein Klick auf den mailto:-Link für Austragungswünsche öffnet nicht nur wie üblich ein Mailer-Fenster mit der korrekten To:-Zeile, sondern trägt auch gleich die Betreffzeile „Wrong Blacklist Entry“ ein. Auf solche Weise lassen sich diverse Voreinstellungen vornehmen, besonders für sogenannte „Role Accounts“ (info@, abuse@ etc.), die häufig unter Spam-Massen begraben werden und dringend auf Positivmerkmale in den eingehenden Mails angewiesen sind.

Vorsicht beim Verwenden von im Web verbreiteten „Mailto-Generatoren“: Viele setzen Vorgaben wie „Vorname Nachname <mail@example.com>“ nicht korrekt um, da sie die spitzen Klammern nicht in HTML-Entities verwandeln. Ein persönlicher, korrekter Name in der Adresszeile ist aber ein besonders starkes Indiz für erwünschte Mails (häufig „Ham“ genannt) und sollte auf jeden Fall auf der eigenen Webseite auftauchen, um Erstkontakte zu unterstützen. Auf diese Weise kann ein einziger E-Mail-Account mehreren Anwendern und unterschiedlichen Zwecken dienen: Je nach „Realname“ lässt sich eine E-Mail in verschiedene Mail-Ordner zustellen, etwa mittels Procmail oder IMAP-Sieve, und ganz ohne Namensangabe vielleicht in den Ordner „Spam-Verdacht“.

Neben dem Abgrasen von Webseiten stehen Spammern viele weitere Methoden zur Adressengewinnung zur Verfügung. Dazu zählt das simple Durchprobieren aller möglichen Adressen (Wörterbuchangriff). Daher sollte man sich für einen zunächst unbequem erscheinenden „Local Part“ – den persönlichen Teil vor dem Domainnamen – in seiner Adresse entscheiden, etwa „nachnavo“ statt „vorname.nachname“. Alles, was in Namens- oder anderen Wörterbüchern auftaucht, finden Spammer nämlich eines Tages, sodass sich der eigene Name schon einmal nicht als Adressbestandteil anbietet. Im Gegenteil: Mit dem fest an die E-Mail-Adresse geknüpften Namen können sich Informationen über Geschlecht, Herkunft und

iX-TRACT

- Spam-Versender steigern ihren Ausstoß laufend, doch schwierig zu filternde „gut gemeinte“, aber unerwünschte E-Mails wie Spam-Empfangsbestätigungen und nicht selbst bestellte Newsletter sind ein mindestens ebenso stark wachsendes Ärgernis.
- E-Mail-Anwender können mit gewissen Vorsichtsmaßnahmen beim Einrichten und Verwenden ihrer Adressen den Spam-Anteil in ihrem Posteingang senken.
- Nicht nur Versender seriöser Werbung, sondern alle Internet-Nutzer tragen Verantwortung dafür, nicht als Spam-Quelle in Erscheinung zu treten.

Mitgliedschaft beenden

Wenn Sie Ihren Account wirklich löschen möchten, geben Sie zur Bestätigung bitte Ihr Passwort hier ein:

Passwort ************

Wenn Sie zusätzlich verhindern möchten, dass Sie unter Ihrer E-Mail-Adresse erneut eingeladen werden, können Sie Ihre Adresse bei uns sperren lassen. Bitte beachten Sie, dass Sie in diesem Fall, auch keinen neuen Stammbaum unter dieser Adresse anlegen können!

Ja, ich möchte meine E-Mail-Adresse bei verwandt.de sperren lassen.

Löschen Abbrechen

Eine Robinsonliste zum Sperren von Adressen gegen ein erneutes missbräuchliches Eintragen hat leider Seltenheitswert (Abb. 1).

sogar das ungefähre Alter in Spammerkreisen verbreiten.

Auch Spammer haben ihre Gewohnheiten, und dazu gehört das Nutzen eines beschränkten Zeichensatzes. Wer Adressen einrichtet, die sonst selten genutzte Zeichen enthalten, etwa Plus- oder Gleichheitszeichen, darf damit rechnen, wenig oder gar keinen Spam zu erhalten.

Zum Glück sind einfach zu merkende E-Mail-Adressen selten wirklich notwendig, denn in mindestens 99 Prozent der Anwendungsfälle setzt sie das Mailprogramm oder der Webbrowser automatisch ein, statt dass sie jemand eintippt. Manchmal sind simple Adressen sogar kontraproduktiv. Im Fall von GMX mit zahlreichen Top-Level-Domains etwa führen vermeintlich wertvolle Adressen wie „john@…“ und „michel@…“ immer wieder zu Verwechslungen und bergen damit die Gefahr von Zustellfehlern oder Irrläufern (siehe Kasten „Was tun mit Mail-Irrläufern?“).

Und das ist nicht die einzige Komplikation: Wer sich eine einfache Freemailer-Adresse à la „paul@gmx“ einrichten möchte, benötigt dafür nicht nur eine gehörige Portion Glück, dass die Wunschadresse frei ist. Er muss sich auch darüber im Klaren sein, dass solche Adressen mit Sicherheit schon mindestens einen Vorbesitzer gehabt haben, wenn der Freemail-Provider den betreffenden Domainnamen bereits jahrelang anbietet. Und wer eine „gebrauchte“ Adresse benutzt, kann es noch lange zu spüren bekommen, was der Vorbesitzer damit alles angestellt hat.

Eine einzige Adresse genügt

Ebenfalls Spam-minimierend wirkt es sich aus, wenn ein Unternehmen nicht sämtliche seiner Domainnamen (example.com/net/org…) für E-Mail nutzt, sondern sich für eine Domain entscheidet, über die es sämtliche E-Mails abwickelt. Die Zeiten, da der Besitz vieler gültiger Mailadressen als schick galt, sind längst vorbei. Adressen, die gar nicht erst im Einsatz sind, können auch keinen Spam anziehen. Wer mehrere TLDs nutzt, tut das entweder in der Hoffnung, mehr Besucher auf seine Website zu locken oder Konkurrenten ein Schnippchen zu schlagen, die ebenfalls an der Domain interessiert sein könnten – nicht aber zu dem Zweck, den eigenen Briefkopf mit Mailadressen zu füllen, die alle zum selben Ziel führen.

Wer beim Einrichten seines Domainnamens noch vor der Auswahl einer Top-Level-Domain (TLD) steht, kann davon ausgehen, dass eine Domain umso mehr Spam anzieht, je länger die TLD auf dem Markt ist. So haben einige vor gut einem Jahr registrierte, sehr einfache *iX*-Testadressen unter der seinerzeit neu eingerichteten Domain gmx.eu bis heute keine einzige Spam-Mail angezogen, während einfache Adressen unter gmx.com und gmx.de ohne Spamfilter kaum mehr benutzbar sind.

Wie viel Sorgfalt auch immer ein Mail-Anwender beim Einrichten seiner Adresse walten lässt: Jede noch so trickreich vor Web-Harvestern und Wörterbuchangriffen geschützte Adresse kann früher oder später dennoch in der Datenbank eines Spammers landen. So sind immer wieder automatische Antworten an Mailinglisten zu beobachten, die in Webarchiven landen. Neben Webseiten Dritter gibt es weitere öffentliche Quellen, etwa Whois-Datenbanken oder Schlüsselserver, in denen sich die Adressen nicht verschleiern lassen.

Besonders interessante Adressquellen für kriminelle Spammer sind „enteignete“, Malware-verseuchte PCs, die sich nicht nur als „Bots“ zum Versenden von Datenmüll, sondern auch zum Ausspähen persönlicher Informationen des ursprünglichen Besitzers nutzen lassen. Dazu gehören neben digitalen Adressbüchern und dem eigentlichen Mail-Bestand praktischerweise auch die E-Mail-Accounts – einer der Gründe dafür, warum auch ganz normale Mailserver seriöser Internetprovider als Spamquellen auftreten können.

Wer halbwegs sicher sein möchte, dass sein PC nicht zum Teil eines Botnetzes wird, sollte die üblichen Maßnahmen gegen PC-Verseuchung ergreifen: Keine Mail-Anhänge aus unbekannter Quelle öffnen, keine Software zweifelhafter Herkunft installieren und einen aktuellen Malware-Scanner einsetzen. Und wenn der PC nur während der Arbeitszeit und nicht rund um die Uhr läuft, spart das nicht nur 75 % der Energie, sondern mindert auch mögliche Verseuchungsfolgen. Dem Missbrauch eigener Ressourcen durch Unbekannte begegnet man nicht zuletzt auch dadurch, kein ungeschütztes WLAN und keinen PC ohne Router direkt am Internet zu betreiben.

Sonder(ab)fall Bounces

Nicht jede unerwünschte Mail im Posteingang ist direkt auf Tätigkeiten krimineller Spammer zurückzuführen. Dazu gehören überraschende Benachrichtigungen über den Zustellstatus von Spam-Mails („Bounces“) und viele andere automatisch erzeugte Nachrichten, die deren Empfänger nicht selbst ausgelöst hat, etwa Abwesenheitshinweise und Empfangsbestätigungen von Unbekannten oder gar Bitten um eine Absender-Verifizierung (zusammenfassend „Backscatter“ oder auch „Collateral Spam“ genannt). Hinzu kommen eigentlich seriös erscheinende Werbemails, die der Besitzer der Empfängeradresse aber niemals bestellt hat. All diesen Erscheinungen ist gemeinsam, dass ein Spamfilter sie schlecht zu fassen bekommt. Hier sind die Verursacher gefragt, Gegenmaßnahmen zu ergreifen.

Die Ursache für Bounces liegt darin, dass Spam häufig gestohlene Absenderadressen (Envelope-From) trägt, an die Fehlermeldungen und andere automatische Antworten gehen. Pauschal unterdrücken sollte man insbesondere keine Bounces mit Unzustellbarkeitsbenachrichtigungen (Non-Delivery Notifications, NDN), denn sie können sich ja durchaus auch auf tatsächlich selbst versendete E-Mails beziehen und gehören zwingend zum SMTP-Standard.

Ein Mailserver sollte möglichst wenige dieser schwer in den Griff zu bekommenden Nachrichten versenden. E-Mail basiert jedoch auf Speichern und Weiterleiten, daher lassen sich Bounces nie ganz vermeiden.

Blacklist-Risiko durch Abpraller

Nicht nur die Rücksichtnahme gebietet es, seine Mailsysteme weitgehend Bounce-frei zu gestalten. Wer viele Bounces generiert, läuft stets Gefahr,

auf schwarzen Listen zu landen und dann nicht mehr nur keinen Backscatter, sondern gar keine Mails mehr versenden zu können. Ähnliches gilt für Urlaubsantworten und Empfangsbestätigungen. Das Versenden automatisch generierter E-Mails birgt stets ein hohes Risiko und sollte nur dann erfolgen, wenn einigermaßen sichergestellt ist, dass die verursachende E-Mail gerade keinen Spam oder Virus enthält. Wer selbst E-Mails nach dem Motto „Von Ihnen ging Spam aus“ versendet, dem ist nicht bekannt, dass die Absenderadressen in Spam nicht authentisch sind und in vielen Fällen Unbeteiligten gehören. Leider gehören dazu sogar einige Hersteller von Anti-Spam-Lösungen, die es ihren Kunden ermöglichen, auf den Eingang von Spam mit einer Antwort-Mail zu reagieren.

Zu den besonders hartnäckigen Verursachern von „Collateral Spam“ gehören die sogenannten Challenge-Response-Systeme (CR). Sie verlagern den Aufwand des Filterns vom Empfänger auf den Absender – und vor allem auf die vielen Besitzer der von Spammern missbrauchten Absenderadressen. Schon wegen des hohen False-Positive-Risikos verbieten sich CR-Lösungen von selbst, denn die Challenge-Mails müssen ihrerseits Spamfilter passieren oder führen gar zu Blacklistings. Auch wenn CR-System A auf CR-System B trifft, kann es nicht mit der Kommunikation klappen.

Ärger mit Massenmails

Ebenfalls massenhaft und automatisch versendet, aber mehr oder weniger persönlich adressiert, befördert das sogenannte „Permission Marketing“ täglich Millionen von Werbemails („Newsletter“) in die Eingangsordner. In vielen Fällen sind es tatsächlich angeforderte Kundeninformationen, doch häufig nehmen es die Absender nicht so genau damit (siehe Kasten). Einen großen Unterschied zum „echten“ Spam sehen viele unfreiwillige Empfänger in den Newslettern nicht, und tatsächlich sind die Übergänge fließend.

Rolf Anweiler vom Newsletter-Dienstleister eCircle schätzt, dass im Konsumentenbereich bis zu 20 Prozent eines Adressenbestandes jährlich nutzlos werden können – sie sind also entweder gar nicht mehr erreichbar oder wechseln den Besitzer. Anweiler bringt den Grund dafür auf den Punkt: „Wenige Kunden überprüfen im E-Mail-Marketing wirklich sorgfältig und vor allem regelmäßig Ihren Adressenbestand, da die Rückläufer im Gegensatz zum Brief nur geringe Kosten verursachen.“ Mit anderen Worten: Das Medium ist sogar dann zu billig, wenn legale Kunden dafür bezahlen. Und daher wird es noch lange E-Mails geben, die niemand haben will und die dennoch Millionen von Empfängern erreichen. (un)

Ein brauchbarer Newsletter und sein Gegenstück

Gelegentlich eingesetzte Positiv-Merkmale	Oft vorkommende Negativbeispiele
From: Beispielfirma <info@example.org>	From: noreply@example.com
To: Bert Ungerer <un@ix.de>	To: Unsere Kunden
Subject: Preissenkungen im Dezember	Subject: Herr asdf, Gratisinfo für Sie
Guten Tag, Herr Ungerer, anbei finden Sie Neuigkeiten aus unserem Programm. (Es folgen diverse gut lesbare Informationen.)	Klicken Sie hier, wenn Ihr Mailprogramm den Inhalt nicht richtig darstellt. (Bild auf einem externen Server, von dem das Mailprogramm aus Datenschutzgründen nur den Umriss anzeigt)
Klicken Sie hier, wenn Sie den Newsletter nicht mehr erhalten wollen, oder senden Sie eine leere E-Mail an optout-un=ix.de@example.org	Sie erhalten diesen Newsletter, weil Sie sich bei uns oder bei unseren Partnern angemeldet haben. Zum Verwalten Ihrer Einstellungen loggen Sie sich bitte ein.

Ein brauchbarer Newsletter geht nur an bestätigte Adressen und gibt darüber Auskunft, an welche Adresse er ging und wie sie sich auf einfache Weise aus der Empfängerliste löschen lässt. Weniger empfängerfreundlich sind Eintragungen durch Dritte, unerreichbare Absender, eine anbiedernde bis aggressive Anrede in der Betreffzeile und komplizierte Austragungsprozeduren.

Literatur

[1] Jochen Topf et al., Antispam-Strategien; Unerwünschte E-Mails erkennen und abwehren; BSI/Bundesanzeiger Verlag, Köln 2005

[2] Jochen Topf; Querschläger; Versand von Bounces minimieren; *iX* 7/2006, S. 150

[3] Verband der deutschen Internetwirtschaft e. V.; eco Richtlinie für zulässiges E-Mail-Marketing; www.eco.de/dokumente/Richtline_OM.pdf

[4] Bert Ungerer; Sendungsbewusst; E-Mail-Marketing: Fehler beim Newsletter-Versand und deren Vermeidung; *iX* 8/2006, S. 92

Werbung auf Abwegen

Viele – auch namhafte – Unternehmen schaffen es auch nach langer Internet-Präsenz nicht, nur verifizierte Empfängeradressen zum Versenden ihrer Newsletter zu verwenden („Double Opt-in“). Unter anderem bei den folgenden Unternehmen, die zum Teil massiv auf Werbung per E-Mail setzen, kann anscheinend jeder jede Adresse eintragen: Apple, Autoscout24, aspect online, Axa, Cinemaxx, Conrad, ino24, einsurance, Procter & Gamble, Skype, Tchibo, Universal Music und Volkswagen.

Auch das laufende Überwachen der verwendeten Adressen auf Erreichbarkeit, das Austragen unerreichbarer oder jahrelang passiver Kunden gelingt selbst großen Unternehmen häufig nicht, was nicht zuletzt Sicherheitsbedenken aufwirft. So erhalten neue Besitzer alter Adressen unter anderem Kundenpost von 1&1, Amazon, Commerzbank, DAB bank, Deutsche Telekom, Ebay, Freenet, GMX, O2, Paypal, Talkline, Tiscali.

Es scheint schwierig zu sein, E-Mail-Werbung nicht an die falschen Adressaten zu richten. Daher sollten darin Informationen enthalten sein, wie sie sich umgehend abstellen lässt, etwa per unmittelbar nutzbarem Web-Link, ohne Login, Passwort, Bestätigungsmails oder andere Komplikationen. Schwierig ist die Austragung jedoch zum Beispiel für unfreiwillige „Kunden“ der Commerz- und der DAB bank, von Freenet oder KLM.

Manche Absender bieten eine Austragung per Antwortmail an, jedoch leider selten unabhängig von der tatsächlichen Adresse, die ja ein Weiterleitungsziel sein kann. Und eine „Robinsonliste“ zwecks Sperrung einer Adresse gegen jegliche zukünftige Aussendung hat Seltenheitswert.

Wenn keiner der genannten Wege funktioniert, bleiben nur noch Beschwerden bei den technisch verantwortlichen Absendern. Mit etwas Glück findet der unfreiwillige Empfänger seinen Kandidaten in der „Certified Senders Alliance“, die der eco-Verband ins Leben gerufen hat (certified-senders.de). Dann kann eine Mail an whitelist-complaints@eco.de helfen, die Fehlerursache zu beheben. In vielen anderen Fällen ist es nicht so leicht, überhaupt jemanden mit einem offenen Ohr für Werbegeplagte zu finden.



IP-Blacklists sinnvoll kombinieren

Blockwerk

Christian Rossow, Christian Dietrich, Norbert Pohlmann

Trotz eines Spam-Anteils jenseits der 90 Prozent funktionieren die Mailserver bei Providern, Unternehmen und anderen größeren Organisationen, und die Anwender nutzen ihr Lieblingsmedium weiterhin unverdrossen. Großen Anteil daran haben IP-Blacklists, von denen viele Postmaster sogar mehrere einsetzen.

Blacklists sind umstritten. Doch solche Verzeichnisse von IP-Adressen Spam-versendender Rechner bilden anerkanntermaßen einen wichtigen Schutzmechanismus im Kampf gegen den Missbrauch von Internet-Ressourcen. Eine Reihe von ihnen lässt sich kostenlos nutzen. Bei der Auswahl von Blacklists zur Spam-Abwehr auf einem Mailsystem verlassen sich viele Systemadministratoren und IT-Entscheider aufs Hörensagen oder auf ihr Gefühl. Die Erfahrung anderer ist – wie so oft – ein wichtiger Gesichtspunkt bei der Auswahl von Blacklists. Das Folgende stellt, basierend auf empirischer Inhaltsanalyse einiger frei verfügbarer Blacklists, weitere Anhaltspunkte dar, die bei der Auswahl und insbesondere beim Kombinieren von Blacklists helfen können.

Das Markieren, mehr noch das Blockieren von Spam anhand der IP-Adresse des Absenders hat den Vorteil, dass es enorme Ressourcen sparen kann. Der SMTP-Dialog wird dann in der Regel bereits in einem frühen Stadium durch das annehmende Mailsystem unter Angabe eines Fehlercodes beendet (Reject). Es findet also keine Übertragung des Inhalts der E-Mail statt, und der annehmende Mailserver muss sich gar nicht erst um die Verarbeitung kümmern. Dies macht sich umso vorteilhafter bemerkbar, je stärker der Mailserver ausgelastet ist. Anti-Spam-Maßnahmen wie Inhalts- und Virenfilter erfordern sehr viel I/O- und Rechenleistung.

Blockieren als zweischneidiges Schwert

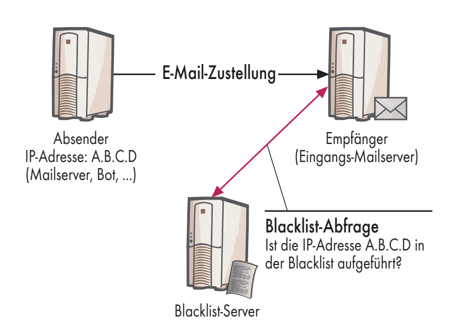
Allerdings ergibt sich ein potenzieller Nachteil beim Einsatz von IP-Blacklisting. Wer beispielsweise – wie es immer wieder vorkommt – einen eigentlich legitimen Ausgangs-Mailserver blockiert, den Spam-Versender missbrauchen, enthält seinen Anwendern auch erwünschte E-Mails von dort vor. Der gelistete Provider gerät dadurch unter Druck und muss sich um eine Austragung seiner IP-Adresse kümmern, will er seine Kunden behalten. Blacklist-Betreiber und -Anwender müssen insbesondere bei großen Providern mit Augenmaß vorgehen, da sonst möglicherweise ein einzelner Benutzer durch den Versand von Spam Tausende von Anwendern desselben Mailservers in Mitleidenschaft ziehen kann. Andererseits hilft dieser Druck mitunter, dass Provider schnell reagieren und den Spammer aus ihrem Netz befördern. Leider gibt es aber auch viele Beispiele, in denen Provider weder auf Beschwerden über spammende Kunden noch auf Blacklistings reagieren.

Es ist daher für Anwender von IP-Blacklisting umso wichtiger, auf eine nachvollziehbare und vertretbare Art und Weise Spam-Abwehr zu betreiben: Zum einen, um den eigenen Kunden und Anwendern einen angemessenen Schutz zu bieten, zum anderen, um im Falle eines Listings eine klare Begründung geben zu können.

Blacklists arbeiten unterschiedlich

Vor diesem Hintergrund ist es interessant, inwieweit die Inhalte, also die gelisteten Adressen verschiedener Blacklists, einander überschneiden oder auch ergänzen. Dies könnte Administratoren Hinweise darauf geben, wie viele und vor allem in welcher Kombination Blacklists sinnvoll zur Spamabwehr geeignet sind.

Blacklists unterscheiden sich hinsichtlich der Policy, die darüber bestimmt, welche IP-Adressen auf die



Mithilfe einer während des SMTP-Dialogs abgefragten, oft von externen Dienstleistern betriebenen IP-Blacklist können Mailserver darüber entscheiden, welche E-Mails sie als spamverdächtig markieren oder gleich ablehnen. Fürs Filtern nach der Zustellung beim Endanwender sind sie weniger gut geeignet (Abb. 1).

Umfang einiger IP-Blacklists

Name	Adressbereich	Einträge	IPv4-Anteil	Nutz-Anteil
pbl.spamhaus.org	320 152 000	130 000	7,4541 %	18,3826 %
xbl.spamhaus.org	5 789 000	5 789 000	0,1348 %	0,3324 %
CBL	5 212 000	5 212 000	0,1214 %	0,2993 %
all.dnsbl.sorbs.net	5 090 000	2 836 000	0,1185 %	0,2923 %
dnsbl.njabl.org	4 459 000	4 459 000	0,1038 %	0,2561 %
dnsbl.ahbl.org	3 488 000	3 132 000	0,0812 %	0,2003 %
dnswl.org (Whitelist)	2 867 000	9 000	0,0668 %	0,1647 %
sbl.spamhaus.org	1 807 000	5 000	0,0421 %	0,1038 %
UCEPROTECT L1	801 000	801 000	0,0187 %	0,0460 %
NiX Spam	78 000	78 000	0,0018 %	0,0045 %

Liste gelangen können. Oft sind Blacklists in mehrere Subzonen mit speziellen Policies unterteilt. Darüber hinaus unterscheidet sich die Dauer des Listings einer IP-Adresse. Manche Blacklists entfernen IP-Adressen nach einer definierten Zeit, beispielsweise die Blacklist der *iX*-Redaktion (www.nixspam.org) nach derzeit vier Tagen. Andere lassen IP-Adressen für unbestimmte Zeit in der Blacklist und entfernen Einträge allenfalls bei begründeten Beschwerden der Gelisteten.

Blacklists weisen unter anderem wegen der unterschiedlichen Policies deutliche Unterschiede in ihrem Umfang auf. Einige führen lediglich einzelne IP-Adressen unabhängig voneinander auf, die beispielsweise nur dann Eingang in die Liste finden, wenn sie tatsächlich als Spam-Quelle auftreten. Andere listen dagegen ganze IP-Adressbereiche, etwa dann, wenn in bestimmten Subnetzen eine gewisse Zahl einzelner IP-Adressen als Spam-Quellen aktiv und ein Kollateralschaden in der „Adressnachbarschaft“ unwahrscheinlich ist. Solche Blacklists können wesentlich größere IP-Adressbereiche abdecken, ohne dass tatsächlich von jeder einzelnen IP-Adresse Spam ausging.

Der Anteil der gelisteten Adressen lässt sich an der theoretischen Gesamtzahl von IP-Adressen messen. IPv4 stellt rund 4,2 Milliarden zur Verfügung. Für die Praxis ist allerdings der Anteil der gelisteten Adressen an der praktisch nutzbaren Menge zugewiesener IP-Adressen aussagekräftiger (siehe Tabelle), die derzeit etwa bei der Hälfte der Gesamtzahl liegt.

Optimieren durch Vergleichen

Eine besondere Stellung nimmt in der Vergleichstabelle die Liste dnswl.org ein. Es handelt sich nicht um eine Black-, sondern um eine Whitelist mit dem Ziel, die Zahl von Fehlalarmen durch gelegentlich auf Blacklists auftauchende „richtige“ Mailserver zu reduzieren. So gibt es eine Menge Spam, der zum Beispiel von Google- oder Yahoo-Accounts ausgeht, aber nicht jeder möchte solche namhaften Mail-Provider aussperren, wenn sie mal wieder die Aufnahme in eine Blacklist ausgelöst haben. dnswl.org listet gezielt solche Server, denen man zutrauen kann, vor allem tatsächlich erwünschte E-Mails zu versenden.

Eine Vergleichsmatrix zeigt den Anteil, der sich in zwei zu vergleichenden Blacklists überschneidet. Die Angabe der Werte erfolgt prozentual, gemessen an der Blacklist der Zeile.

Beim Interpretieren der vorliegenden Auswertung ist zu beachten, dass es sich bei der Analyse von Blacklists stets um eine Momentaufnahme handelt. Sie ändern sich durch zum Teil sehr schnelles Hinzufügen und Entfernen von Adressen laufend. Die Zahlen der hier aufgeführten Blacklists stammen vom 12. Juli 2007. Nicht alle Blacklist-Betreiber bieten von sich aus ihren kompletten Datenbestand zum Download an. Bei einigen war viel Überzeugungsarbeit dafür notwendig, ausnahmsweise Zugriff auf die Listen zu erhalten.

Die Listen in Form von DNS-Zonen oder als Text-Downloads komplett anzubieten, ergibt aber aus mehreren Gründen Sinn. Einerseits motiviert man Vielnutzer, die Listen zu spiegeln und dadurch sowohl viel Traffic zu sparen als auch die Informationen über Kommunikationspartner zurückzuhalten, die sie sonst mit den DNS-Anfragen preisgeben. Zum anderen sind die Nutzer weniger abhängig vom Anbieter, falls dessen eigene DNS-Server ausfallen.

Im Vergleich zeigen sich deutlich die Beziehungen der Blacklists von Spamhaus untereinander. Offensichtlich umfasst die XBL (xbl.spamhaus.org) die CBL komplett. Außerdem enthält die PBL einen Großteil der CBL (circa 74 %). Darüber hinaus deckt die PBL große Teile anderer Blacklists ab. So sind beispielsweise IP-Adressen der von der *iX* gepflegten Blacklist NiX Spam zu mehr als 55 % in der Liste von Spamhaus enthalten. Umgekehrt führte NiX Spam zum Zeitpunkt der Erhebung mit 0,007 % nur einen kleinen Bruchteil der PBL, was wegen des enormen Umfangs der PBL immerhin rund 21 600 Einträge ausmachte.

iX-TRACT

- Zum Entlasten von Mailservern, Verbessern der Filterquote und um Druck auf spammerfreundliche Provider auszuüben, setzen viele Postmaster Blacklists mit IP-Adressen ein, von denen Spam ausgeht.
- Diverse Blacklists bieten sich für den Einsatz an, doch nicht jede Kombination eignet sich für einen effizienten Mailserver-Betrieb.
- Selbst umfangreiche Blacklists decken weniger als 10 Prozent des gesamten IPv4-Adressbereichs ab und treffen über den Rest keine Aussage.
- Zukünftig dürften Verfahren an Bedeutung gewinnen, die allen IP-Adressen eine mehr oder weniger gute Reputation zuordnen können.

Einige IP-Blacklists im Vergleich (gemeinsame Anteile in %)

Blacklist	all.dnsbl.sorbs.net	UCEPROTECT L1	NiX Spam	dnsbl.ahbl.org	sbl.spamhaus.org	dnsbl.njabl.org	CBL	pbl.spamhaus.org	xbl.spamhaus.org	dnswl.org	Bogus ranges
all.dnsbl.sorbs.net	–	1,83	0,28	10,17	10,67	11,03	8,03	36,92	17,92	0,002	7,73
UCEPROTECT L1	11,61	–	2,34	1,97	0,58	2,93	64,14	69,96	64,79	0,026	0,01
NiX Spam	18,32	23,80	–	1,79	0,64	2,58	41,02	55,36	42,58	0,064	0,02
dnsbl.ahbl.org	14,83	0,45	0,04	–	0,56	64,32	3,74	66,38	13,87	0,002	0,22
sbl.spamhaus.org	29,15	0,25	0,03	1,04	–	0,88	1,23	5,49	1,49	0,003	9,68
dnsbl.njabl.org	12,59	0,53	0,05	50,31	0,37	–	4,75	67,11	21,03	0,003	0,28
CBL	7,84	9,86	0,62	2,50	0,44	4,07	–	73,91	100,00	0,001	0,00
pbl.spamhaus.org	0,58	0,17	0,01	0,72	0,03	0,93	1,19	–	1,36	0,000	1,48
xbl.spamhaus.org	15,39	8,76	0,57	8,17	0,47	15,83	88,05	73,92	–	0,001	0,01
dnswl.org (Whitelist)	0,003	0,007	0,002	0,003	0,002	0,005	0,001	0,002	0,002	–	0,027
Bogus ranges	0,03	0,00	0,00	0,00	0,01	0,00	0,00	0,34	0,00	0,000	–

Beim Kombinieren von Blacklists spielen die jeweiligen Schnittmengen für die Gesamtwirksamkeit eine große Rolle. Die Tabelle zeigt den prozentualen Anteil an IP-Adressen der Blacklist A (Zeile), den auch Blacklist B (Spalte) abdeckt.

Die PBL umfasst alle von Providern als „Mailserver-frei“ gemeldeten Adressen, die in der Regel dynamisch an Einwahl- und DSL-Kunden vergeben werden, also nicht nur verifizierte Spamquellen.

Es zeigt auch auf, welche Kombinationen von Blacklist-Abfragen sinnvoll sind. Zwei Blacklists mit geringen Überschneidungen zu vereinen, ist wesentlich effektiver, als solche zu kombinieren, die eine hohe Anzahl von IP-Adressen übereinstimmend aufführen. So erscheint es wenig sinnvoll, Spamhaus’ XBL mit der CBL zu kombinieren, die in der XBL enthalten ist. Andererseits lässt sich die NiX-Spam-Liste gut zusammen mit den Listen von SORBS einsetzen, da mehr als vier von fünf Einträgen der NiX-Spam-Liste nicht von SORBS gelistet sind.

Abbildung 2 zeigt den gesamten Adressraum von IPv4. Die obere Hälfte der Grafik repräsentiert die NiX-Spam-Blacklist, der untere Teil stellt den Inhalt der PBL grafisch dar. Die Horizontale ist in 256 Adressblöcke mit jeweils 16 777 216 (2^{24}) IP-Adressen unterteilt. In der CIDR-Schreibweise entspricht dies sogenannten /8-Netzen. Die Y-Achse gibt logarithmisch an, wie viele IP-Adressen in der jeweiligen Blacklist (oben NiX Spam, unten PBL) gelistet sind. Die Länge der senkrechten Linien verdeutlicht, dass es zwar hinsichtlich der Anzahl der Einträge Unterschiede gibt, dass jedoch kaum Adressblöcke nur in einer der beiden Listen vertreten sind.

Black- versus Whitelists

Whitelists eignen sich hervorragend zum Schutz vor falschen Einträgen in Blacklists. Bei auf Whitelists gelisteten IP-Adressen handelt es sich meistens um legitime E-Mail-Server. Von ihnen ausgehende E-Mails sollen von Black- oder Greylisting unbehelligt bleiben. Da aber auch gewöhnliche Mailserver gelegentlich zum Versenden von Spam missbraucht werden, können durchaus manche IP-Adressen sowohl auf Blacklists als auch auf Whitelists gelistet sein.

Wie bereits erläutert, kann ein einziger Anwender eines E-Mail-Systems den Ruf des gesamten Systems und somit all seiner Nutzer schädigen. Im schlimmsten Fall führt dies zur Eintragung des Servers in Blacklists. Solche Systeme kann man durch Whitelisting schützen und dadurch das verbindungsorientierte Blacklisting umgehen. Deshalb ist es sinnvoll, neben Blacklisting andere Filter einzusetzen, darunter Inhaltsfilter, die eine Entscheidung pro individueller E-Mail treffen und auch beim Aussetzen von Blacklisting noch Spam filtern können. Die Vergleichsmatrix zeigt, dass sich die

CBL: Zuordnung zu autonomen Systemen

AS	Bezeichnung	Einträge	lokaler Anteil [%]
9121	TTNET	234 145	2,393
4134	Chinanet-Backbone	205 120	0,298
27699	Telecom. de Sao Paulo	157 017	6,347
7738	Telecom. da Bahia	152 159	3,034
8167	Telecom. de Santa Catarina	148 816	6,894
8151	Uninet S. A. de C. V.	135 221	1,101
4837	China169-Backbone CNC Group	112 558	0,531
9829	BSNL National Internet Backbone	106 412	1,527
7643	Vietnam Posts and Telecom.	101 770	10,492
3269	Telecom Italia	99 629	0,826

dnsbl.njabl.org: Zuordnung zu autonomen Systemen

AS	Bezeichnung	Einträge	lokaler Anteil [%]
4134	Chinanet-Backbone	708 818	1,031
4837	China169-Backbone CNC Group	285 048	1,344
4766	Korea Telecom	193 308	0,782
27699	Telecom. de Sao Paulo	164 522	6,650
9318	Hanaro Telecom	161 925	1,849
7132	AT&T Internet Services	126 151	0,404
7738	Telecom. da Bahia	112 192	2,237
22927	Telefonica de Argentina	102 464	13,192
3462	HINET Data Comm.	84 589	1,094
8151	Uninet S. A. de C. V.	76 139	0,620

Überschneidungen zwischen diversen Blacklists und der Whitelist dnswl.org mit Werten unterhalb des Promillebereichs in Grenzen halten.

Verbotene Zonen im Internet

Ebenfalls interessant ist der Vergleich von Whitelists mit sogenannten Bogons, nicht zugewiesenen und damit eigentlich nicht aktiven Netzbereichen. Sie umfassen neben per Definition nicht benutzbaren oder reservierten Netzadressen auch von der IANA noch nicht vergebene IP-Adressen. Externer IP-Verkehr aus diesen Bereichen kann also ohne Auswirkungen auf die Erreichbarkeit komplett und schon am Router blockiert werden. Für Blacklists hat es somit keine negativen Auswirkungen, falls sie Bogons umfassen. Gänzlich anders ist die Ausgangssituation bei Whitelists: Für Bogons ist eine Listung zum Umgehen von Blacklisting kontraproduktiv.

Beim Blick auf die Vergleichsmatrix fällt auf, dass die Whitelist dnswl.org eine gewisse Überschneidung mit den als Bogon deklarierten Netzadressen aufweist. Auf Anfrage hat der Betreiber der Whitelist diese Einträge inzwischen gelöscht. Es handelte sich wahrscheinlich einfach um Tippfehler. Der Betreiber will zukünftige Einträge auf Bogons prüfen, um neue derartige Fehleinträge zu vermeiden.

Fehlerhafte Listungen einer Whitelist wie dieser können fatale Folgen haben. Schafft es ein Spammer, von einer IP-Adresse auf der Whitelist aus E-Mails zu versenden, verzichten je nach Bekanntheits- und Nutzungsgrad der Whitelist viele Eingangs-Mailserver auf ein Blacklisting. Das kann zur starken Mehrbelastung nicht nur des Servers, sondern auch der Empfänger führen, zu denen dann mehr Spam durchdringt.

Vergleich mit Routing-Informationen

Das Analysieren von Blacklists erlaubt auf bequeme Weise eine Zuordnung von unangenehm auffallenden IP-Adressen zu Ursprungsländern oder autonomen Systemen (AS). So lassen sich Statistiken über die Provider und Länder mit den meisten Spam-Quellen erstellen, ohne auch nur eine einzige Spam-Mail selbst empfangen zu müssen. Hier dienten die Blacklists CBL und dnsbl.njabl.org diesem Zweck. Die CBL enthielt zum Zeitpunkt der Analyse etwa 5,2 Millionen, dnsbl.njabl.org umfasste rund 4,5 Millionen IP-Adressen.

Ein autonomes System ist eine Einheit mit gemeinsamen Routing-Informationen – etwas vereinfachend, aber meist korrekt mit „Internetprovider“ beschrieben. Die zehn am häufigsten von Blacklistings getroffenen AS sind aufgeführt. Es fällt zum Beispiel auf, dass sich das chinesische AS 4134 auf beiden Blacklists weit oben befindet. Bei den meisten handelt es sich um asiatische oder amerikanische AS, aber auch die europäische Telecom Italia ist in der Listung der CBL zu finden. Die Spalte rechts gibt an, welchen Anteil des jeweiligen AS die Blacklist enthält. So ragen beispielsweise das AS 7643 auf der CBL mit 10,5 % und das AS 22927 auf dnsbl.njabl.org mit 13,2 % des gesamten für das AS verfügbaren Netzbereichs hervor.

Eine etwas abstraktere Sicht ermöglicht die Zuordnung von IP-Adressen zum Herkunftsland. So kann man teilweise die regionale Ausrichtung einer Blacklist erkennen und schwarze Schafe unter den Ländern ausmachen. Negativ fallen in beiden Blacklists die Länder China, Brasilien und die Vereinigten Staaten auf. Aber auch europäische Länder gibt es innerhalb der „Worst 10“: Deutschland und Polen sind – wenig schmeichelhaft – oft in der CBL vertreten, Frankreich hat viele Einträge in dnsbl.njabl.org. Darüber hinaus fällt in der NJABL-Liste der hohe Anteil (3,9 %) an gelisteten IP-Adressen im Verhältnis zu Argentiniens gesamtem verfügbaren Adressraum auf. Das als ergiebige Spam-Quelle bekannte China ist in absoluten Zahlen tatsächlich jeweils

dnsbl.njabl.org: Zuordnung zu Staaten

Land	Einträge	Anteil dort [%]
China	1268571	0,977
USA	476130	0,019
Korea	448338	0,772
Brasilien	441364	1,881
Indien	194092	1,412
Argentinien	186102	3,884
Frankreich	131657	0,099
Mexiko	118733	0,612
Taiwan	113741	0,607
Japan	96239	0,043

CBL: Zuordnung zu Staaten

Land	Einträge	Anteil dort [%]
Brasilien	563159	2,400
China	424435	0,327
USA	359369	0,014
Türkei	235375	2,867
Russland	223684	1,391
Indien	219334	1,596
Deutschland	196971	0,283
Korea	196485	0,339
Mexiko	152294	0,785
Polen	142674	1,169



Eine grafische Gegenüberstellung der PBL (unten) mit NiX Spam macht deutlich, dass Spammer oft ähnliche Adressblöcke zum Versenden der Nachrichten verwenden (Abb. 2).

recht weit oben angesiedelt. Im Vergleich zu anderen Ländern sind dort jedoch relativ kleine Teile des verfügbaren Adressraums gelistet.

Globale IP-Reputationsdatenbank

Angesichts der zum Teil stark voneinander abweichenden Inhalte diverser Blacklists stellt sich die Frage, welcher Anteil des insgesamt im Internet genutzten IPv4-Adressraums sich durch die Vereinigung aller Blacklists überhaupt abdecken ließe. Es stellt sich heraus, dass die genannten Black- und Whitelists zusammen gerade einmal zu 20 Prozent aller nutzbaren rund 2 Mrd. IP-Adressen überhaupt eine Aussage treffen. Der theoretische IPv4-Adressraum beläuft sich sogar auf rund 4,2 Mrd. IP-Adressen, die also von Black- und Whitelists heutzutage nur zu etwa 10 % abgedeckt sind.

Umgekehrt zeigt dies leider, dass Spammer durch die Nutzung von „frischen“, ungelisteten IP-Adressen in mindestens 80 % aller Fälle nicht von einer Blacklist gesperrt werden können. Es verdeutlicht aber auch das hohe Potenzial künftiger IP-Reputationssysteme. (un)

CHRISTIAN ROSSOW

ist Mitarbeiter im Forschungsbereich E-Mail-Sicherheit des Instituts für Internet-Sicherheit an der FH Gelsenkirchen.

CHRISTIAN DIETRICH

ist Mitarbeiter des Instituts für Internet-Sicherheit an der FH Gelsenkirchen und für den Forschungsbereich E-Mail-Sicherheit verantwortlich.

PROF. DR. NORBERT POHLMANN

ist Informatikprofessor für Verteilte Systeme und Informationssicherheit sowie Leiter des Instituts für Internet-Sicherheit an der FH Gelsenkirchen.

Literatur

[1] Bert Ungerer; Eingeschränkt; IP-Blacklists gegen unerwünschten Datenverkehr; *iX* 4/2007, S. 102

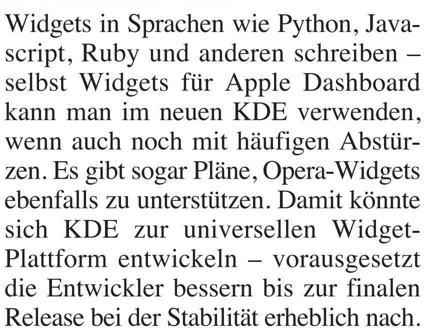
[2] Manuel Schmitt; Entlastungsfrage; Echtzeit-DNS-Blacklist als Mittel gegen Spam; *iX* 12/2007, S. 112

Ein erster Blick auf KDE 4

Speiender Drache

Markus Franz

Im ständigen Wettstreit um die bessere grafische Arbeitsumgebung haben die KDE-Entwickler eine neue Runde eingeläutet. Für die Version 4.0 polierten sie nicht nur die Optik, sondern schraubten auch kräftig unter der Haube. Ein Blick auf den neuen Desktop und die Technik im Hintergrund.



Nach vielen Monaten emsiger Programmierarbeit steht die Version 4.0 in den Startlöchern: Am 20. November gaben die KDE-Entwickler den Release Candidate 1 ihrer Desktop-Umgebung frei. Für die Version 4.0 krempelte die Community die grafische Umgebung für Unix- und Linux-Rechner kräftig um. KDE erfreut sich nach wie vor großer Beliebtheit, alle wichtigen Distributionen liefern es standardmäßig mit aus oder bieten das Nachrüsten an. Obwohl es in den letzten zwei Jahren Marktanteile gegen Gnome verloren hat (vor allem durch Ubuntu), läuft auf vielen Unternehmens-Desktops noch immer ein gepflegtes KDE.



Mit der neuen Release wollten die Entwickler viele angestaubte Bibliotheken komplett neu schreiben und technisch zu Mac OS X und Windows Vista aufschließen. Eine Liste der Neuigkeiten findet sich auf der Homepage des Projekts (siehe „Onlinequellen“ [a]). Natürlich spiegelt sich in einigen Punkten auch die Konkurrenz zum Gnome-Projekt wider. Es fällt bei KDE 4.0 sofort auf, dass die Änderungen nicht nur an der Oberfläche, sondern auch unter der Haube erfolgt sind: Es gibt eine ganze Reihe neuer Frameworks. Zum einen soll es damit einfacher sein, neue Programme für KDE zu schreiben oder existierende anzupassen. Zum anderen sollen sie KDE gleichzeitig beschleunigen und benutzerfreundlicher machen.

Dabei sind die im Folgenden ausführlicher behandelten neuen Frameworks keine Konkurrenz zu existierenden Tools: Wie der Kasten „Neue Komponenten in KDE 4.0“ zeigt, ziehen sie sozusagen nur eine weitere Abstraktionsschicht zwischen den alten Tools und Frameworks sowie der Benutzeroberfläche mit Programmen und Einstellungsmanagern ein.

Oberfläche mit Plasma

Beim ersten Start von KDE 4.0 fällt dem Benutzer sofort die überarbeitete Oberfläche auf. Sie wirkt klarer und strukturierter als das bisherige Pendant. Mit Plasma [b] haben die KDE-Entwickler einen Ersatz für KDesktop, Kicker und andere Kernkomponenten der Oberfläche entworfen. Plasma stellt zum ersten Mal eine integrierte KDE-API für Entwickler bereit, um vom kleinen Applet/Widget zur großen Anwendung schnell Ergebnisse zu erzielen. Mit dem Scripting-Framework Kross lassen sich Widgets in Sprachen wie Python, Javascript, Ruby und anderen schreiben – selbst Widgets für Apple Dashboard kann man im neuen KDE verwenden, wenn auch noch mit häufigen Abstürzen. Es gibt sogar Pläne, Opera-Widgets ebenfalls zu unterstützen. Damit könnte sich KDE zur universellen Widget-Plattform entwickeln – vorausgesetzt die Entwickler bessern bis zur finalen Release bei der Stabilität erheblich nach.

Die „Plasmoids“ genannten Programme bezeichnen sowohl externe Widgets als auch interne KDE-Tools wie den CPU-Monitor. Die Grundlage eines Plasmoids bildet eine Komponentenarchitektur, die aus drei Teilen besteht:
DataSource: Hierunter versteht man eine einheitliche Schnittstelle der Anwendungen zu irgendwelchen externen Daten (auch Systeminformationen wie CPU-Auslastung oder Speicherbelegung).
DataEngine: Dieser Teil bezeichnet die Anwendungslogik eines KDE-4.0-Plasmoids. Er bereitet die mit der Data-Source angezapften Informationen für die Visualisierung auf.
Visualisierung: KTrader erledigt das Laden und Anzeigen aller Plasmoids.

Diese Trennung der verschiedenen Schichten eines Plasmoids ermöglicht es sozusagen, das Model-View-Controller-Schema für Widgets beziehungsweise Desktop-Applets zu verwenden.

Fans der „alten“, für Superkaramba in KDE 3.5 geschriebenen Widgets, kommen ebenfalls auf ihre Kosten: Zur Sicherung der Rückwärtskompatibilität

iX-TRACT

- Mit Plasma, Solid, Phonon und Decibel führt KDE 4.0 neue Frameworks ein.
- Für die Suche setzt der Open-Source-Desktop Strigi mit semantischen Funktionen ein.
- Neben Linux läuft KDE 4.0 auch unter BSD, Mac OS X und Solaris.



Fast alles neu: Die KDE-Entwickler überarbeiteten nicht nur das Look & Feel, sondern auch die darunterliegenden Abstraktionsschichten gründlich (Abb. 1).

integrierten die Entwickler Superkaramba als eine Teilbibliothek in Plasma, sodass alte Widgets ohne Probleme laufen sollten. Dies wollte in der Praxis jedoch nicht gelingen. Man darf gespannt sein, welche Änderungen die Final-Release an bestehenden Applikationen erfordert, damit alles wie versprochen funktioniert.

Ein Vorteil des Gnome-Projekts war es bisher, dass es viele Elemente der Oberfläche konsequent nach Richtlinien zur Barrierefreiheit entwickelt hat. Hier schließt neben einigen Anpassungen an Plasma das Oxygen-Projekt [c] auf: Es definiert ein neues Icon-Thema für KDE 4.0, das den Standards von FreeDesktop.org folgt. Damit garantiert es nicht nur Barrierefreiheit, sondern auch ein einheitliches Aussehen der K-Anwendungen über alle Grenzen hinweg.

Neuer Dateiverwalter

Bisher war der Konqueror sowohl fürs Dateimanagement als auch fürs Surfen im Web zuständig. Die erste Aufgabe übernimmt jetzt standardmäßig Dolphin (auch wenn Konqueror weiterhin Dateien verwalten kann). Mit dieser Neuentwicklung gab das KDE-Team der jahrelangen Kritik nach, dass durch die beiden Einsatzbereiche der Konqueror nur noch schwer zu bedienen sei und einfache Dateioperationen unnötig viel Aufwand erforderten. Tatsächlich geht per Dolphin die Arbeit mit Dateien schneller von der Hand.

Konqueror wurde zu einem reinen Webbrowser degradiert. Seine Oberfläche wirkt wesentlich klarer und die Menüs übersichtlicher. Durch die Radikalkur rendert er durchweg alle Websites nun schneller als zuvor – auch dank der neuen Rendering-Engine Webkit (früher: KHTML). Ins Auge springt der neue Download-Manager, der sich nun endlich direkt in die Symbolleiste integriert.

Im Umgang mit Dateien hilft nun anstelle von KPDF der Betrachter Okular, der mit beinahe jedem Medientyp zurechtkommt. Er beherrscht sogar Windows-Hilfedateien in den Formaten CHM (Compiled HTML Help) und XPS (XML Paper Specification). Anwender können jetzt Anmerkungen zu Dokumenten dauerhaft speichern – besonders praktisch bei PDF-Dateien. Der Bildbetrachter Gwenview erfuhr nur marginale Veränderungen: Die Menüs sind entschlackt und der Benutzer kann Bilder nun beschneiden.

Um die Änderungen rund um den Umgang mit Dateien abzuschließen, noch ein Satz zum Editor Kate: Er ist nun Grundlage für die Editorkomponenten von Konqueror, KEdit und weiteren Applikationen. Das neue Plug-in-System und die verbesserte Suchfunktion kommen so vielen Anwendungen zugute.

Jeder KDE-Anwender hat sicher die eine oder andere unangenehme Begegnung mit dem HAL-Daemon gehabt. Er war für das Einbinden von Geräten und Treibern zuständig. Gleichzeitig gab es mit dem KNetworkmanager ein Tool speziell für (WLAN-)Netze, das manchmal ebenfalls ein instabiles Eigenleben entwickelte.

Solid begreift Geräte

In KDE 4.0 erweitert das Solid-Framework [d] diese beiden Komponenten: Es stellt eine zentrale API für alle Geräte und Dateisysteme bereit. Als Basis dient weiterhin HAL, der also nicht verschwindet – aber für Anwendungsentwickler und Benutzer komplett in den Hintergrund tritt.

Solid entwarfen die Entwickler vor allem dazu, die Unterstützung für im laufenden Betrieb hinzugefügte oder entfernte Geräte zu verbessern. Intern ist die API in viele themenspezifische Domains unterteilt, die sich um ein bestimmtes Thema intensiv kümmern: Ein Bereich kümmert sich um Netze und abstrahiert den Zugriff auf den Networkmanager, sodass jede KDE-Anwendung hiermit interagieren kann. Das stabilisiert in KDE die Verwaltung von kabellosen Netzen und VPNs – die Konfiguration sieht in der Oberfläche aber weitgehend identisch aus. Sogar KPPP gibt es noch in KDE 4.0.

Eine weitere Domain von Solid behandelt das Power Management: So kann der Nutzer (oder eigene KDE-Anwendungen) hier an vielen Optionen drehen – und im Hintergrund läuft wieder HAL. Freunde dürfte der neue Bluetooth-Stack in Solid finden, der endlich diese Funktion im KDE-Kern nachrüstet: Das Suchen und Verbinden zu Geräten funktionierte im Test schon problemlos. Bei Dateiübertragungen traten aber noch Abstürze auf.

Zu guter Letzt sollte man sich das Kommandozeilenprogramm *solidshell* ansehen. Über den Befehl *solidshell network set wireless diabled* lässt sich das WLAN schnell deaktivieren. Mit diesem Tool hat man ein mächtiges Werkzeug an der Hand, das richtig Spaß macht.

Multimedia bündeln

Die Wiedergabe von Multimedia hat schon manchem Linux-Anwender den Schweiß auf die Stirn getrieben: Es fehlt ein Codec hier, dort irgendein Plug-in. Damit will – ähnlich der Intention von Plasma und Solid – ein neues Framework aufräumen: Phonon [e]. Es befreit KDE von GStreamer und Xine und ersetzt den altgedienten Soundserver *aRts*. Die Grundidee hinter Phonon ist es, die gemeinsamen Funktionen aller Medienabspieler – Wiedergabe von Dateien, Start, Pause, Stop – in einem Framework zu bündeln. Die zentralen Einstellungen dazu nimmt der Anwender im sogenannten „Phonon Settings Manager“ vor.

Angelehnt an die Domains bei Solid unterteilt sich Phonon in sogenannte Engines, die beispielsweise für Play, Pause und Suche zuständig sind. Im Zusammenspiel mit Solid bietet Phonon in KDE 4.0 eine bessere Unterstützung von Headsets und anderen Audiogeräten – im Test zeigte sich tatsächlich auf Notebooks eine bessere Steuerung und Stabilität. Insbesondere das integrierte Mikrofon eines Vaio-Notebooks wird nun nicht nur durch Linux mit dem passenden Treiber versorgt, sondern lässt sich auch endlich verwenden.

Ähnlich wie Solid basiert das Phonon-Framework selbst auf den „alten“ Tools wie GStreamer und dem Soundserver. Der Vorteil liegt aber darin, dass die Entwickler bei API-Änderungen an diesen Low-Level-Frameworks nicht alle

KDE-Programme, sondern lediglich Phonon anpassen müssen. Phonon kann laut Dokumentation sogar während des Betriebes das Basis-Framework (beispielsweise von GStreamer zu Xine) wechseln – ohne Unterbrechung im Betrieb.

Es spricht Decibel

Zu den zahlreichen neuen Frameworks gesellt sich noch Decibel [f] dazu. Anders als bei den Brüdern und Schwestern für Desktop, Geräte und Multimedia handelt es sich nicht um einen erweiterten Aufsatz für bestehende Dienste. Vielmehr ist es eine vollkommen neue Basisarchitektur für jegliche Kommunikation: Es bietet eine API mit Schnittstellen für Protokolle jeder Art. In der Theorie bedeutet dies vor allem, dass man in KDE ab sofort ein zentrales Adressbuch hat. Möchte man nun beispielsweise mit einer Person telefonisch in Kontakt treten, startet die passende Anwendung – egal ob Skype, Instant Messenger oder der VoIP-Client.

In der Praxis merkt man von dieser eigentlich genialen Idee noch nichts: Nicht einmal Kopete oder KMail, die Standardclients für Messaging und Mail, greifen auf das neue Framework zu. Die Entwickler haben angekündigt, dass dies auch im finalen KDE 4.0 noch nicht oder nur teilweise der Fall sein wird – erst mit KDE 4.1 sollen die K-Anwendungen vollständig an Decibel andocken.

Mit Strigi schneller suchen

Für die KDE-3-Familie diente Beagle mit einem KDE-Frontend als Suchwerkzeug. Ab KDE 4.0 übernimmt Strigi diese Aufgabe. Das Projekt ist zwar noch recht jung, aber integriert sich hervorragend in die KDE-Bibliotheken. Außerdem ist Strigi bedeutend schneller als Beagle: Bei Tests auf einer mit rund 950 MByte Dateien verschiedenster Typen gefüllten externen Festplatte war Strigi etwa 40 % schneller als der Indizierungsvorgang mit Beagle. Das liegt zum einen daran, dass Strigi von Beginn an auf Geschwindigkeit optimiert wurde. Zum anderen verwendet es – im Gegensatz zu Beagle – ein komplett eigenes System, genannt Jstream, zur Tiefenindizierung des Dateisystems. Neben der üblichen Unterstützung gängiger Formate (Open Document, PDF, MP3, Plain Text et cetera) versteht sich Strigi auch mit Debian- und RPM-Paketen. Mithilfe eines SHA-1-Hash findet es zuverlässig Duplikate im Suchindex.

Zusammen mit dem Informations-Framework Nepomuk bildet Strigi die Grundlage des „semantischen KDE-Desktop“. Dieser soll alle Programme und Dateien über eine einheitliche, integrierte Suchplattform zur Verfügung stellen. Bis auf eine normale Desktopsuche merkt man in der aktuellen Vorabversion aber von diesem engagierten Konzept nichts. Geplant ist, dass in kommenden Versionen Nepomuk eine Basis für Ontologien und beliebige Metadaten bildet.

Eine Frage der Plattform

KDE 4.0 fühlt sich unter Linux auf den ersten Klick schneller an als der Vorgänger. Das liegt nicht nur an den komplett neuen, effizienter gestalteten Bibliotheken, sondern vor allem an Qt 4.0 – dem zentralen Framework, auf dem KDE traditionell basiert. Die 3er-Versionen von KDE nutzen die Qt-Features nicht so effizient aus wie die neue Hauptversion 4.0. Wem Geschwindigkeit weniger wichtig ist, der kann KDE 4 auch auf *BSD, Solaris oder Mac OS X verwenden. Für Mac OS X 10.4 und 10.5 gibt es im Web eine Kurzanleitung [g]. Dort findet man ein wöchentlich aktualisiertes DMG-Paket, das alles beinhaltet, was Anwender für KDE auf dem Mac benötigen. Die Performance dürfte aber deutlich schlechter sein als unter Linux.

Abschließend sei noch auf die KDE Development Platform hingewiesen. Hierin bündeln die Entwickler des KDE-Teams die wichtigen Bibliotheken, um eigene Applikationen zu entwickeln. Dazu gehören APIs für PIM-Anwendungen, I/O-Operationen, die Kern-APIs von KDE und der Notification Daemon, um mit KDE und anderen K-Programmen zu kommunizieren.

Fazit

Schon der aktuelle Release Candidate von KDE 4.0 lässt Freude aufkommen: Es macht richtig Spaß, auf diesem Desktop zu arbeiten. Die neue Version zeichnet sich aber weniger durch die Anpassung der Oberfläche als vielmehr durch die neuen APIs aus: Sie machen KDE schneller, stabiler und effizienter für Entwickler. An der einen oder anderen Fensterkante sollte das KDE-Team noch einmal putzen. Dann kann KDE aber in der finalen Version (die laut Roadmap schon mit Erscheinen dieser Ausgabe freigegeben sein soll) mit Mac OS X, Windows Vista und Gnome auf Augenhöhe um die Gunst des Nutzers buhlen – nicht nur optisch, sondern auch technisch. (avr)

MARKUS FRANZ

ist Mitgründer und CTO der BF Blogform Search GmbH in Berlin und studiert Wirtschaftsinformatik an der Friedrich-Schiller-Universität in Jena.

Neue Komponenten in KDE 4.0

Decibel	gemeinsames Adressbuch aller K-Anwendungen und einheitliches Kommunikations-Interface
Dolphin	Dateimanager, Ersatz für Konqueror
Oxygen	benutzerfreundliche, einheitliche Icons in allen K-Anwendungen
Phonon	Multimedia-Framework für Audio, Video und zugehörige Endgeräte
Plasma	neue Oberfläche für KDE und neue Bibliotheken
Solid	Framework-API zum einheitlichen Zugriff auf Geräte
Strigi	Desktop-Suche mit semantischen Funktionen

Onlinequellen

[a]	KDE 4.0	techbase.kde.org/Schedules/KDE4/4.0_Release_Schedule
[b]	Plasma	plasma.kde.org
[c]	Oxygen	www.oxygen-icons.org
[d]	Solid	solid.kde.org
[e]	Phonon	phonon.kde.org
[f]	Decibel	decibel.kde.org
[g]	KDE auf Mac OS X	ranger.users.finkproject.org/kde/index.php/Home



Apples iPhone als Geschäftstelefon

Kleine Erleuchtung

Christian Kirsch

Apple peilt mit seinem iPhone bislang vor allem Privatkunden an. Viele von ihnen dürfte jedoch der Preis abschrecken. In Unternehmen wiederum steht der Funktionsumfang im Vordergrund. Lässt sich der Flachmann auch dort verwenden?

Seit Monaten ist die Aufregung um Apples iPhone groß. Fast könnte man glauben, es sei die Quadratur des Kreises gelungen. Schon der erste Blick auf das Gerät enthüllt den wesentlichen Unterschied zur Konkurrenz: Es ist nur eine einzige Taste zu sehen. Apple vertraut bei der Bedienung seines Handys ganz auf den großen Touchscreen (320 × 480 Pixel). Anders als Palm OS- und Windows-Mobile-Modelle ist er ausschließlich mit dem Finger bedienbar, auf Stifte oder anderes lebloses Zeigegerät reagiert er nicht. Folglich sieht der Bildschirm schon nach kurzem Benutzen aus, als sei dem Anwender das Buttermesser ausgerutscht.

Nimmt man die Werbung für das Gerät als Anhaltspunkt, sieht Apple das iPhone vor allem als Lifestyle-Attribut. Im Folgenden soll jedoch interessieren, ob es sich auch für den geschäftlichen Einsatz eignet. Die Anforderungen definieren einerseits Konkurrenten wie Blackberrys Curve und Nokias E61i. Andererseits gibt es für reisende Geschäftsleute Unverzichtbares: E-Mail gehört ebenso dazu wie der Zugriff aufs Web und auf Unternehmensdaten. Microsofts Office-Formate sollten zumindest zu betrachten sein; für lokales Bearbeiten gäbe es Bonuspunkte.

Die üblichen PIM-Anwendungen (Personal Information Management) bringt das iPhone mit – allerdings fehlt die Aufgabenverwaltung. Terminplaner, sowie Adressverwaltung sind vorhanden und synchronisieren ihre Daten via USB-Kabel und iTunes mit den Anwendungen auf dem Desktop. Auf einem Mac sind dies das von Apple gelieferte Adressbuch und *iCal*, unter Windows kann man Adressen mit Outlook (Express) und Yahoo abgleichen. Für Termine steht dort nur das Synchronisieren mit Outlook zur Verfügung. In iTunes sind die zu synchronisierenden Kalender und Kontaktgruppen einstellbar. Das mitgelieferte Google Maps und der Mailclient greifen auf die lokal gespeicherten Adressen zu. Dazu reicht es, einen Teil der Mailadresse oder des Namens einzutippen, das Telefon vervollständigt wie üblich die Daten geeignet oder zeigt die Kontaktliste zur Auswahl an.

Länderliste mit lästigen Lücken

In typischer Apple-Tradition nimmt das iPhone seiner Besitzerin möglichst viel lästige Arbeit ab. So muss man das Heimatland bei den Kontaktdaten nicht eintippen, sondern wählt es aus einer Liste aus, bei der Deutschland voreingestellt ist. Dadurch erhält Google Maps immer eine vollständige Anschrift als Eingabe. Es sei denn, der Kontakt wohnt in Thailand, Malaysia oder Laos, denn Südostasien besteht auf dem iPhone nur aus Indonesien und Singapur. In Europa fehlen unter anderem Estland, Lettland, Litauen; Lateinamerika schnurrt auf Argentinien, Brasilien und Mexiko zusammen und Afrika auf Ägypten sowie Südafrika. Warum wohl dürfen iPhone-Nutzer keine Geschäftskontakte in Chile, Marokko oder Ecuador haben – zählen sie zu den Bush'schen Schurkenstaaten? Die Auslassungen wären nicht weiter dramatisch, könnte man ein fehlendes Land manuell ergänzen. Das jedoch geht nicht. Schlimmer: Eine Adresse in Bangkok verlegt das Gerät in der „Bearbeiten"-Sicht nach Deutschland, und Google Maps findet sie nicht.

Termine lassen sich zwar mit einer Notiz versehen, nicht jedoch mit einer

iX-TRACT

- Apples iPhone verwendet nur eine mechanische Taste. Alle anderen Bedienelemente stellt der berührungsempfindliche Bildschirm nach Bedarf bereit.
- Webseiten und Karten lassen sich durch Gesten oder Doppelklick vergrößern und verkleinern. Eine virtuelle Tastatur blendet das Gerät wenn nötig ein.
- Bislang kann man auf dem Telefon keine Anwendungen installieren. Was über die mitgelieferten Programme hinausgeht, muss im Browser laufen.

Liste von Teilnehmern. Wer sie auf dem Desktop erstellt hat, muss damit zurechtkommen, dass sie auf dem iPhone nicht vorliegt: Ein kurzer Anruf etwa durch Anklicken des Namens, um eine Verspätung anzukündigen, ist nur über die Kontaktliste und mithilfe des eigenen Gedächtnisses möglich. Teilnehmer an Treffen sind nicht das Einzige, was beim Datenabgleich zwischen Mac und iPhone abhandenkommt. Ein aus dem Web geladenes und einem Kontakt zugeordnetes Bild übernahm das Telefon trotz mehrerer Versuche nicht. Der umgekehrte Weg, jemanden mit der Handy-Kamera zu fotografieren und seinen Kontaktdaten zuzuordnen, aktualisierte die Desktop-Daten zuverlässig.

Weder Suche noch Filter für Mail

Mail-Kunden von Google, .Mac, AOL und Yahoo finden Einstellungsdialoge vor, in die sie nur Namen, Mailadresse und Passwort eingeben müssen. Für Geschäftskunden dürfte das jedoch kaum relevant sein. Ihnen stehen frei konfigurierbare POP3-, IMAP- und Exchange-Konten zur Verfügung. Im Test funktionierte die IMAP/SSL-Verbindung ebenso anstandslos wie ein Exchange-Konto. Letzteres benutzte die freie z-Push-Software (z-push.sourceforge.net), die IMAP-Server via Activesync so ins Netz bringt, dass sie wie Exchange-Server wirken. Die Mail-Konfiguration beschränkte sich in allen Fällen auf die Angabe des Server- und Benutzernamens samt Passwort sowie der Mailadresse. Dass die IMAP-, SMTP- und Exchange-Zugänge durch SSL/TLS gesichert waren, merkte das iPhone selbstständig, sodass Benutzer weder Ports noch Verschlüsselungsverfahren eintragen mussten. Dies ist nur bei vom Standard abweichenden Einstellungen nötig.

Auf Wunsch ruft das iPhone Mails regelmäßig ab oder überlässt es seiner Eigentümerin, dies manuell zu erledigen. Filtermöglichkeiten sind jedoch nicht vorgesehen, sodass sie zwangsweise auch unterwegs alles zu sehen bekommt, was auf dem Mailserver eintrudelt. Ebenso überwältigt sind IMAP-Anwender bei den Foldern: Weder lässt sich ihre Hierarchie auf- und zuklappen noch die Anzeige auf die abonnierten Ordner beschränken. Schwerer wiegt das Fehlen von Suche und Sortierung: Wer viele Nachrichten auf dem iPhone gespeichert hat, muss sie in Eingangsreihenfolge durchblättern, bis er die richtige findet. Word-, Excel- und PDF-Anhänge zeigt der Mailclient weitgehend korrekt an, mit Powerpoint und einfachen Textdateien kann er jedoch nichts anfangen. Da es auch für Office-Dokumente keine Funktionen zum Bearbeiten oder Suchen gibt, sind größere Excel-Tabellen nicht zweckmäßig zu nutzen. Eine von Apples eigenem *iCal*-Kalender verschickte Einladung zu einem Termin kann der Mailclient übrigens weder öffnen noch lässt sie sich auf dem iPhone in den dortigen Terminkalender übernehmen.

Dreh- und Angelpunkt für jede zusätzliche Applikation auf dem iPhone ist der mitgelieferte Browser, der auf dem aus dem KDE-Projekt stammenden Webkit basiert. Bislang nämlich setzt Apple ganz auf das Web 2.0 und

auf per Ajax aufgehübschte Webanwendungen. Erst im Februar soll ein SDK herauskommen, das es ermöglicht, native Anwendungen für das iPhone zu erstellen. Der mitgelieferte Safari bietet alle wesentlichen Fähigkeiten des ausgewachsenen. Statt seiner Tabs verwendet die Miniversion virtuelle Seiten, zwischen denen der Anwender durch zwei Klicks wechseln kann. Nicht mehr benötigte Seiten schließt er durch Antippen der roten Markierung an ihrer linken oberen Ecke.

Javascript und CSS bilden keine spürbaren Hürden für den Browser, SVG und Flash jedoch verarbeitet er nicht. Webseiten lassen sich auf dem Handy nicht speichern, man muss sie also erneut laden, wenn man sie wiederum benötigt. Dass ausgerechnet der Erfinder von Cut & Paste keinen Mechanismus bereitstellt, um Inhalte von Webseiten zu markieren, zu kopieren und anderswo einzufügen, enttäuscht. So ist es in der Regel nicht möglich, eine Nummer aus einer HTML-Seite direkt zum Wählen zu verwenden. Das funktioniert nur, wenn sie als Link mit dem wenig gebräuchlichen *tel:*-„Protokoll“ formatiert ist. Übrigens fehlen nicht nur Cut, Copy und Paste, sondern auch Undo. Deshalb lässt sich ein Fehler nicht einfach reparieren.

Beim Tempo hängt der iPhone-Browser die Konkurrenz ab. Per WLAN dauerte das komplette Laden von www.heise.de/newsticker rund acht Sekunden, via Edge gut 20. Zum Vergleich: Nokias E61i benötigte mit WLAN etwa 26 Sekunden, mit UMTS noch 33. Ähnlich sah das Ergebnis bei Seiten aus, die aus vielen Komponenten bestehen. Offenbar kommt es also für das Tempo beim Browsen nicht nur auf die jeweilige Verbindung an – wichtig sind auch die Qualität des Browsers und die Leistungsfähigkeit der CPU. Damit verliert einer der Hauptkritikpunkte am iPhone an Bedeutung. Dass es bislang nicht per UMTS, sondern nur via EDGE funkt, hielten viele hierzulande für einen schwerwiegenden Nachteil. In der Praxis dürfte er nur beim Versenden oder Herunterladen großer Dateien ins Gewicht fallen. Da das iPhone eine solche Funktion jedoch ohnehin nicht bietet, sollte die langsamere Anbindung nicht stören.

Unangenehm fällt auf, dass dem iPhone das Äquivalent zum „Schlüsselbund“ des Mac-Desktop fehlt. Diese Anwendung speichert auf Wunsch Anmeldedaten, Zertifikate et cetera und erspart im Safari das Eintragen von Benutzernamen und Passwort zur Authentifizierung bei Websites. iPhone-Anwender müssen auf diesen Komfort verzichten. Als einziger lokaler Speicher für Daten steht ihnen die Notiz-Anwendung zur Verfügung, die jedoch alle Daten im Klartext speichert – deshalb ungeeignet als Passwort-Safe.

Bislang bietet das iPhone keine VPN-Verbindung via IPSec. Zumindest mit PPTP und L2TP lässt sich jedoch eine verschlüsselte Verbindung in andere Netze herstellen. Beide Protokolle arbeiten mit Passwörtern oder RSA-Schlüsseln. Im Test funktionierte der PPTP-Client anstandslos.

Neben WLAN und GPRS/EDGE steht als dritte Funktechnik Bluetooth zur Verfügung. Es dient jedoch ausschließlich zum Anschließen von Headsets und Freisprecheinrichtungen. Deshalb ebnet das Mobiltelefon weder einem Notebook den Weg ins Internet, noch kann man mal eben schnell Kontaktdaten, Termine oder gar Dateien zwischen ihm und einem anderen Gerät austauschen. Apple zwingt damit die Anwender, jede Visitenkarte abzutippen – da war Palm OS vor Jahren schon weiter.

Bekehrung der Ungläubigen

Betrachtet man nur die vom iPhone angebotenen Funktionen, scheint der Einsatz für Geschäftszwecke zumindest eingeschränkt möglich: Es kommt mit mehreren Mailkonten zurecht, bietet eine Art VPN, der Browser zeigt die üblichen Webseiten korrekt an – das ist mehr, als man zum Beispiel von Blackberry-Geräten sagen kann. Dass man bislang weder eigene Dokumente noch andere Software auf dem Gerät speichern kann, schränkt den Nutzen jedoch stark ein – nicht nur für den geschäftlichen Bereich. Profile, die auf üblichen Handys verschiedene Einstellungen gruppieren (etwa für laute und leise Umgebungen), fehlen dem iPhone. Dadurch muss man lästigerweise jede einzelne Lärmquelle separat umschalten.

Umgekehrt gilt jedoch: Andere Geräte bieten dasselbe wie das iPhone, teilweise sogar mehr, da sich fehlende Funktionen mit Programmen von Drittanbietern nachrüsten lassen. Trotzdem löst Apples Hardware bei vielen einen sofortigen starken „Haben-wollen“-Reflex aus. Das liegt weniger an der puristischen Form – ausschlaggebend ist die Bedienung, bei der sogar gestandene Handy-Agnostikern zu Gläubigen werden. Hört man Erstnutzern zu, ist immer wieder „So hätte es schon immer sein müssen“ zu vernehmen. Wenn ein Mobiltelefon selbst eine Botschaft versenden kann, dann das iPhone: „function follows form“.

Aller Anfang ist leicht gemacht

Form bedeutet in diesem Zusammenhang in erster Linie „Bedienung“. Kollegen und Freunde, die das iPhone während des Tests in die Hand bekamen, fanden sich darauf ohne weitere Erklärungen zurecht. Darauf verlässt sich auch die mitgelieferte Dokumentation, die aus einem knappen Leporello besteht: Sie erklärt nur die ersten Schritte mit dem Telefon, alle Details muss der Anwender durch Versuch und Irrtum herausfinden. Das funktioniert in der Regel schnell und ohne Frustrationen.

Webseiten und Straßenkarten vergrößert man durch Auseinanderziehen zweier Finger oder durch zweimaliges Antippen. Statt eines Rollbalkens bewegt wiederum der Finger den sichtbaren Bereich direkt. Virtuelle Schieber übernehmen das An- und Ausschalten von Funktionen. Ebenfalls virtuell ist die Tastatur, die immer erscheint, wenn der Benutzer ein Textfeld aktiviert. Sie bringt eine automatische Rechtschreibkorrektur mit, die jeweils das nächste (aus ihrer Sicht) passende Wort vorschlägt. Ein Druck auf die Leertaste akzeptiert es, Antippen des Vorschlags lehnt ihn ab. Das erfordert den ständigen Blick auf den eingetippten Text, da das Wörterbuch gelegentlich merkwürdige Korrekturen vorschlägt. So mag es „SMS“ nicht und beharrt stattdessen auf „ANS“. Wer nach dem abschließenden „S“ ohne Kontrollblick die Leertaste betätigt, muss vier Zeichen einzeln löschen. Sogar mehrfaches Korrigieren des iPhone-Vorschlags hilft dagegen nicht: Es beharrte im Test noch nach der 24. korrekten Eingabe von „SMS“ auf „ANS“.

Trotz dieser Unzulänglichkeiten im Detail bleibt der Gesamteindruck vorherrschend, dass Apple bei der Bedienung dieses Mobiltelefons die Konkurrenz meilenweit hinter sich gelassen

hat. Die Anwendungen sind weitgehend sinnvoll integriert. So startet beim Antippen einer Anschrift im Kontakt Google Maps, das Berühren der Telefonnummer wählt sie und ein Klick auf die Mailadresse öffnet eine neue Nachricht. Beim Drehen des Geräts rotiert der Bildschirminhalt mit, sodass man leicht zwischen Hoch- und Queransicht wechseln kann – allerdings nicht bei eingeblendeter Tastatur. Die grau/blau gehaltene Oberfläche lässt alle Elemente gut erkennen; wichtige wie „Senden“ oder „Öffnen“ sind jeweils in einem leuchtenden Blau hervorgehoben. Per Finger lässt sich deutlich schneller navigieren als mit einem Joystick oder einer 4-Wege-Wippe, vor allem auf einer Webseite, wenn es ums Verschieben geht.

Fazit

Funktional kann das iPhone in weiten Bereichen mit anderen Smartphones etwa von Blackberry oder Nokia mithalten. Seine Beschränkung auf EDGE spielt in der Praxis bislang keine Rolle, da es zum einen hinreichend flott arbeitet und zum anderen den Versand beziehungsweise Empfang großer Dateien, bei dem sich eine schnellere Verbindung auszahlen würde, ohnehin nicht gestattet. Überlegen zeigt es sich bei der Oberfläche, die ein wesentlich komfortableres und schnelleres Bedienen ermöglicht als die Konkurrenz. Im Detail gibt es jedoch einige Mängel: Webseiten und andere Dokumente lassen sich nicht speichern, Mails und Kontakte nicht durchsuchen, eine Undo-Funktion fehlt ebenso wie Kopieren und Einfügen. Größtes Manko ist aber der von Apple über native Anwendungen verhängte Bann. Wird er aufgehoben, sodass Fremdhersteller Software für fehlende Funktionen liefern und Unternehmen ihre Anwendungen auf das Gerät bringen dürfen, könnte es seinen Platz im Geschäftsumfeld erobern. (ck)

Lieferumfang und Preise

Hersteller	Apple, www.apple.de
Produkt	iPhone
Maße	115 × 61 × 11,6 (L/B/H mm^3)
Gewicht	135 g
Speicher	8 GByte
Display	320 × 480 Pixel
Funk	GSM, EDGE; WLAN; Bluetooth nur für Headsets/Freisprecheinrichtungen
Akkulaufzeit	Standby 280 h, Sprechzeit 8 h (lt. Hersteller)
Preis ohne Vertrag	861 €

iX-Wertung

⊕ komfortable und intuitiv zu bedienende Oberfläche

⊕ Anwendungen gut integriert

⊕ trotz EDGE schneller Browser

⊖ Beschränkung auf mitgelieferte und Webanwendungen

⊖ keine Profile

⊖ kein Cut & Paste

⊖ kein Suchen in Mail und Kontakten möglich



Mac OS X 10.5 „Leopard“

Neues Fell

Thomas Kaiser

Mit angeblich durch das iPhone verursachter sechsmonatiger Verspätung erschien Apples Mac OS X 10.5, Codename „Leopard“, fast genau zweieinhalb Jahre nach seinem Vorgänger. An der Oberfläche und unter der Haube hat sich einiges getan.

Die neue Version kommt anders als beispielsweise Microsofts Vista nur in einer einzigen, quasi der „Ultimate“-Variante und als Universal Binary, das auf allen PPC- und Intel-Macs funktioniert, die die Mindestvoraussetzungen erfüllen. Diese sind in Form von Intel-CPU oder G5/G4 mit mindestens 867 MHz Taktung, 512 MByte RAM und 9 GByte Plattenplatz so gesetzt, dass Leopoard prinzipiell auf allen Apple-Rechnern der letzten drei bis vier Jahre läuft (Pro-Modelle bis zu 6 Jahre zurück). Um richtig Spaß zu haben, sollte man diese Werte jedoch mindestens verdoppeln, zumal 10.5 viele Fähigkeiten aktueller Grafikkarten nutzt, die auf älteren Modellen fehlen oder nur träge funktionieren.

Das Aktualisieren von der Vorversion läuft wie von Apple gewohnt zumeist glatt – siehe Kasten „Upgrade-Pfad zum Leopard“. Eine Stolperfalle können systemnahe Software und „Haxies“ bilden, konkret Unsanitys Application Enhancer, der Upgrade-Installationen zu Fall bringt. Kritisch ist die Verfügbarkeit von Treibern, die man im Vorfeld prüfen sollte (bezüglich Druckern hilft eine Websuche nach „apple 306684“ beim Finden des Support-Artikels). Wie gehabt, kann man durch „Anpassen“ der Installation und Verzicht auf nicht benötigte Druckertreiber sowie Sprachen ein paar GByte auf der Platte sparen.

Visuell beruhigt, dass die ausufernde Vielfalt der Fensterstile der Vorversion [1] auf nur noch eine etwas biederanmutende Version schrumpft. Zum Ausgleich erregen sich die Gemüter vor allem wegen kaum mehr zu unterscheidender Ordner-Icons sowie verstärkter Transparenzeffekte in Menüs und Menüleiste. Zumindest bei Letzterem ist einfache Abhilfe möglich (siehe Kasten „Tiger-Look“). Das Dock kommt im neuen abschaltbaren gläsernen Look daher und beherbergt eine Novität, die Stacks (Stapel, siehe Abb. 1).

Zieht man Ordner ins Dock, erscheinen sie dort als Stapel. Ein Klick darauf und die im Ordner enthaltenen Dateien klappen inklusive Vorschau nach oben. Enthält er allerdings mehr als circa ein Dutzend Elemente, kommt bei dieser Darstellung die Übersicht abhanden. Weiterer Nachteil: Hierarchische Ordner via Dock sind nicht mehr möglich (Abhilfe siehe Kasten „Tiger-Look“).

Ebenfalls runderneuert präsentiert sich der Finder. Als vierte Darstellungsvariante beherrscht er die von iTunes bekannte Cover-Flow-Ansicht, die vor allem im Zusammenspiel mit Spotlight praktisch ist (Abb. 1). Änderungen an lokalen Dateisystemen erscheinen blitzschnell in Finder-Fenstern, und insgesamt wirkt er weniger träge. Die Seitenleiste hat Apple aufgewertet, indem sie standardmäßig (und abschaltbar) typische Spotlight-Suchen einblendet. Gleichfalls tauchen dort per Bonjour im Netz verfügbare Macs auf. Ein Klick auf einen davon genügt und man kann direkt Dateifreigaben nutzen oder den Bildschirm à la VNC übernehmen – passende Zugangsdaten und die Aktivierung von beidem auf dem anderen Mac vorausgesetzt.

Finder zeigt Vorschaubilder an

Eine weitere Neuerung ist Quick Look („Übersicht“), das im Finder schnell, aber trotzdem qualitativ hochwertig den Inhalt von Dateien anzeigt, ohne die zugehörige Anwendung öffnen zu müssen. Proprietäre Dateitypen benö-

iX-TRACT

- Neben offensichtlichen Änderungen an der Oberfläche von Mac OS X 10.5 (Leopard) wie den virtuellen Desktops und einer weiteren Ansicht im Finder gab es zahlreiche Neuerungen im Unterbau.
- HFS+ nutzt wesentlich mehr Metadaten, Bonjour übernimmt das Annoncieren aller Dienste, und Spotlight kann via LAN suchen.
- Das mitgelieferte Backup-Programm „Time Machine“ sichert regelmäßig alle Änderungen auf externen Platten, patzt jedoch noch beim Zurückspielen.

tigen dafür ein passendes Plug-in vom Hersteller oder einem Drittanbieter (eine Liste entsteht gerade unter apfelquak.de/ql-plugins). Für die verbreiteten Dateitypen bringt Leopard passende Module mit.

Nachdem Apple seit fast zweieinhalb Jahren Mehrtastenmäuse verkauft, folgt als weiterer Meilenstein „Spaces“: Ein Desktop-Manager ab Werk, der bis zu 16 virtuelle Bildschirme verwaltet. Offensichtlich traut man den Anwendern inzwischen einiges zu. Das Umschalten zwischen einzelnen Bildschirmen erfolgt per Tastaturkürzel oder automatisch beim Wechsel ins jeweilige Programm. Per Aufruf der Spaces-Anwendung kann man aus der Vogelperspektive Fenster zwischen verkleinerten Darstellungen der Desktops verschieben. Außerdem lassen sich Programme immer demselben Bildschirm zuweisen.

iChat transportiert Bilder, Filme und PDFs

Apples mit Fokus auf Heimnutzer entwickelte Anwendungen bringen in den 10.5 beigelegten Versionen einige Detailverbesserungen. Im professionellen Umfeld dürfte interessieren, dass sich iCal3 dank vollwertiger CalDAV-Unterstützung zur gruppentauglichen Kalenderanwendung entwickelt. Angebunden an einen passenden Server (beispielsweise iCal Server oder sein Open-Source-Pendant Darwin Calendar Server) lassen sich damit in Teams Termine und Ressourcen planen.



Leopards Finder bringt eine neue Seitenleiste und die von iTunes entlehnte Cover-Flow-Darstellung. Datei-Stapel wachsen aus dem vollverspiegelten Dock heraus (Abb. 1).

Apples Instant-Messaging-Programm *iChat* bekam neben allerlei multimedialen Aufhübschungen die Möglichkeit mitgegeben, Präsentationen via iChat-Theater beim Gesprächspartner abzuspielen. In hoher Qualität kann es Bilder, Filme und PDFs zeigen – beziehungsweise jedes Dateiformat, für das es ein Quick-Look-Plug-in gibt. Gleiches gilt für die Bildschirmausgabe von Cocoa-Programmen – Neuübersetzen für 10.5 mit der *IMAVManager*-Klasse und schon erledigt iChat die Ausgabe als Videostream.

Wie im Finder kann man ab Leopard auch in iChat Bildschirmfreigaben so nutzen, dass Partner den eigenen Rechner fernsteuern dürfen und andersherum. Zum Einsatz kommt dabei die aus Apple Remote Desktop (ARD) stammende VNC-Komponente, die wie bei ARD seit Version 3.2 deutlich flotter zu Werke geht. Solange die Firewall nicht zu restriktiv konfiguriert ist, funktioniert das Ganze auch, wenn beide Gesprächspartner hinter NAT-Routern sitzen.

Netzadministratoren sollten diese Art der unfreiwilligen LAN-Öffnung im Auge behalten, wie auch interne Nutzer von Apples kostenpflichtigem .Mac-Service. Ab 10.5 erlaubt es die „Back to my Mac“-Funktion .Mac-Kunden, mehrere Rechner zu durchsuchen und komplett fernzusteuern. Ein reiner NAT-Router, der eingehende Verbindungen blockt, hilft dagegen nicht mehr.

Mail.app in Version 3.1 kommt mit einem Bündel HTML-Vorlagen und versteht sich auf das Abonnieren von RSS-Feeds, die es in Postfächer einsortieren kann. Es erkennt in Grenzen selbstständig Adressen, Telefonnummern und Termine in Nachrichten, die es auch extrahiert und in die passende Anwendung überträgt. Außerdem erlaubt es das Verwalten sowie Bearbeiten von Notizen und Aufgaben. Safari 3 bringt neben besseren Suchfunktionen und der Möglichkeit, Teile von HTML-Seiten als *Dashboard*-Widgets zu speichern endlich die Fä-

Upgrade-Pfad zum Leopard

Hat man noch einen Mac mit OS X 10.2 oder älter, lässt sich ein Upgrade nicht problemlos durchführen. Stattdessen ist die Neuinstallation von Betriebsystem samt Anwendungen und manueller Transfer der Benutzerdaten ratsam.

Ab Mac OS X 10.3 aufwärts sollte man als ersten Schritt einen aktuellen, vollständigen und idealerweise bootfähigen Klon der zu aktualisierenden Platte auf einer weiteren anlegen. Das geht beispielsweise per *SuperDuper!, Carbon Copy Cloner* in Version 3 oder mit Apples Festplatten-Dienstprogramm im „Wiederherstellen“-Modus – für Letzteres siehe Apples Knowledge-Base-Artikel mit ID 306516 (Websuche nach „apple 306516“). Auch sollte man die alte System-Version mit allen verfügbaren Updates auf den aktuellen Stand bringen.

Ist die Integrität des Klons geprüft, kann sowohl eine Aktualisierung des Betriebsystems folgen (hierbei vorher unbedingt eine Dateisystem-Reparatur laufen lassen) als auch das komplette Neuinstallieren samt Migration von Benutzer-Accounts, Einstellungen und Daten per Migrationsassistent vom vorher erzeugten Klon. Beide Varianten eliminieren relativ zuverlässig unnötige Altlasten und übernehmen ein Maximum an Daten und Einstellungen, sodass meist keinerlei weitere Anpassung mehr nötig ist.

Die Migrationsvariante schlägt sich geringfügig besser und hat zugleich den Vorteil, dass sie durch Neu-Initialisierung des Dateisystems ein klassisches Problem von Systemaktualisierungen vermeidet: Die Vergrößerung eventuell vorhandener Inkonsistenzen des Dateisystems, ausgelöst durch die massive Aktivität während der Aktualisierung. Idealerweise lässt man den Migrationsassistenten direkt am Ende der Installation laufen, da er so alle Benutzer-Accounts mit korrekten UIDs migriert und keine Nachbearbeitung mehr nötig ist.

Die Installationsvariante „Archivieren und Installieren“ ist nicht als Ersatz für einen vorher angelegten Klon zu verstehen, weil dies die alte und neue Systemversion auf einem Laufwerk mischt (für Details siehe Article-ID 107120 bei Apple). Das Festplatten-Dienstprogramm wird idealerweise vom Menü der Leopard-DVD gestartet, da es dann Vollzugriff auf die betroffene Platte hat.



Time Machine ist Leopards Backup-Programm, das automatisch regelmäßige Datensicherungen auf externen Festplatten erstellt (Abb. 2).

higkeit, PDFs sinnvoll im Browser zu betrachten.

Zu den nicht offensichtlichen Neuerungen gehört, dass sich Mac OS X ab 10.5 mit Konformität zur Single Unix Specification (SUSv3) und POSIX 1003.1 schmücken darf. Der Darwin-Kernel ist nun bei Version 9.1.0 angekommen, stellt weiterhin eine Mach-BSD-Melange dar und weist neben verbessertem Scheduling, Multi-Core-Optimierung, dynamischem Paging und dynamisch setzbaren Resource Limits zwei echte Neuigkeiten auf: Die von TrustedBSD inspirierte Implementierung der „Mandatory Access Control“ – siehe *sandbox(7)* respektive */usr/share/sandbox/* – ist die eine, die Portierung von Solaris’ *DTrace*-Funktion in den Darwin-Kernel die andere.

dtrace verfolgt Spuren im Programm

Nicht nur Kernel-Hacker freuen sich, dass sie mit *dtrace(1)* und *dtruss(1)* jederzeit an Prozesse und den Kernel andocken und tiefe Einblicke in deren Aktionen gewinnen können. Ab *XCode 3.0* steht in Form von *Instruments,* vormals als XRay bekannt, ein grafisches Frontend dafür zur Verfügung. Fans des ebenfalls von Sun stammenden ZFS (Zeta Filesystem), dürften hingegen enttäuscht sein. Es steht offiziell nur zum Lesen zur Verfügung und ist noch ein Stück davon entfernt, zum Standard- oder auch nur gleichberechtigten Mac-Dateisystem aufzusteigen.

10.5 wertet HFS+ an zwei Stellen auf: Einerseits durch die Implementierung von Verzeichnis-Hardlinks für Time Machine (zum Glück ohne *ln(1)* dieselbe Fähigkeit zu spendieren) und zum anderen durch den stark ausgeweiteten Gebrauch von Metadaten. Es existieren jetzt neben den klassischen Mac-Metadaten, die es schon vor OS X gab, und den seit 10.0 vorhandenen extended Flags von BSD seit 10.4 die Extended Attributes (EA), die Finder und System seit Leopard heftig nutzen.

Zu allem Überfluss wird auch noch dynamisch zwischen den verschiedenen Sets gemappt. Einem klassischen *locked* entspricht das BSD-Flag *user immutable*, dito einem Finder-Kommentar das EA *com.apple.FinderInfo* und vice versa. Einen Überblick über die diversen Flags verschaffen *GetFileInfo(1)* aus *XCode* und das *ls(1)*-Kommando (*-O* für BSD extended Flags, -@ für EA, *-e* ACLs). ACLs (Access Control Lists) setzt Apple seit 10.5 konsequent ein, um einige Standardverzeichnisse vor versehentlichem Löschen oder Bewegen zu schützen. Dagegen durchsetzen kann man sich mit *chmod(1)*, analog hilft für EAs *xattr*, das auch deren Anzeige übernimmt.

Seit Leopard lassen sich auch mit der Desktop-Version von Mac OS X Dateisysteme oder Teile von ihnen in den Systemeinstellungen per AFP, SMB und FTP mit fein granulierten POSIX-Berechtigungen freigeben (ACLs wirken

Zeitmaschine

Time Machine (TM), Apples Leopard beigelegte Backup-Anwendung für Einzelplätze, ist prinzipiell mit einem Klick konfiguriert. Sobald man eine dem System unbekannte Platte anschließt, fragt es automatisch, ob es sie für Sicherungen benutzen soll. Anschließend läuft das Backup unauffällig im Hintergrund – inklusive Versionierung, Rotation der Sicherungen und Benachrichtigung nur bei Bedarf (siehe Abb. 2). Alte Sicherungen löscht TM nach einem Regelwerk bei zu wenig freiem Platz auf dem Medium automatisch. Die Installations-DVD bringt ein Werkzeug mit, das das System aus einem beliebigen vorherigen TM-Backup wieder herstellen soll: Desaster Recovery inklusive.

Unter der Haube bedient sich der *backupd* des FSEvents-Framework, das Änderungen seit der letzten Sicherung registriert. Es findet nur bei Bedarf eine – dann durchaus langwierige – Komplettsicherung statt. Sonst speichert TM lediglich die Änderungen seit dem letzten Lauf. Die Sicherungs-Sets sind auf den ersten Blick ganz normale und in sich vollständige Ordner. Apple arbeitet mit Hardlinks auch auf Verzeichnisse, die alles aus vorherigen Sicherungen einbinden, was sich nicht geändert hat.

Wen die Aussicht beunruhigt, zum Lesen und Schreiben eingehängte Platten zu sichern, der kann sich aufgrund restriktiv gesetzter ACLs für die Backup-Verzeichnisse ein wenig entspannen und außerdem Medien-, also Plattenrotation betreiben. Time Machine merkt sich den Backup-Stand je Medium, sodass man problemlos mit wechselnden Backup-Sets arbeiten kann, um die Gefahr der Datenkorruption zu reduzieren.

Übers Netz funktioniert Time Machine übrigens bislang nicht unmittelbar: Apple hatte die Funktion in Vorversionen von Leopard zwar vorgesehen, hat sie in der endgültigen jedoch abgeschaltet. Im Internet kursieren zwar Tipps, wie man trotzdem SMB- oder NFS-Dateisysteme mit Time Machine nutzen kann. Diese setzen jedoch auf ein dort erzeugtes Plattenabbild mit HFS+-Filesystem, das fürs Backup beziehungsweise Restore zu mounten ist. Mangels Unterstützung durch Apple verbietet sich dieses Verfahren in Produktivumgebungen.

Die Zeitmaschine verdankt ihrem Namen in erster Linie dem Restore-GUI, das eine virtuelle Zeitleiste einblendet. Je nach Verfügbarkeit von Sicherungen auf ihr kann man sich weiter in die Vergangenheit zurückhangeln und auf ältere Versionen von Objekten zurückgreifen. Time Machine arbeitet dabei nicht ausschließlich mit einem Dateimodell. Vielmehr nutzt es bei Anwendungen, die bereits die TM-API verwenden, deren Objektmodell für die Anzeige älterer Versionen.

Beispielsweise kann man im Adressbuch direkt nach Kontakten suchen ohne wissen zu müssen, wo auf der Platte sie liegen. Analog in Mail-Programmen, wo die Spotlight-Integration in TM das Suchen in nur im Backup verfügbaren Datenbeständen ermöglicht. Es steht zu hoffen, dass viele Programme diese Integration mit Time Machine nutzen. Die Komplexität für den Entwickler ähnelt der beim Schreiben von *MDImporter*-Plug-ins für Spotlight.

Zurzeit ist Time Machine als alleiniges Backup-Werkzeug allerdings noch nicht zuverlässig genug. Diverse Anwenderberichte belegen mangelnde Stabilität sowohl beim Backup als auch beim Restore. Und solange Letzterer nicht immer und unter egal welchen Bedingungen klappt, ist ein Backup nichts wert. Es heißt also noch abzuwarten.

trotzdem). Bonjour annonciert alle Dienste, sodass das Auffinden von Ressourcen im Netz komfortabler denn je verläuft. Dasselbe gilt auch für die Druckerfreigabe, deren Basis nun CUPS 1.3 ist, das authentifiziertes Drucken per Kerberos beherrscht.

Apples GUI unabhängig von der Auflösung

Der TCP/IP-Stack von 10.5 justiert optimale Buffergrößen selbstständig abhängig von den Umgebungsbedingungen. Spotlight bekam zudem Netzfähigkeiten spendiert: Passende Berechtigungen vorausgesetzt, kann man auf anderen Leoparden im LAN nach Dateien suchen. Die neue integrierte anwendungsorientierte Firewall erlaubt prinzipiell feiner abgestufte Regeln als der reine Paketfilter der Vorgängerversionen, steht aber schon wegen offensichtlicher Lücken in der Kritik (www.heise.de/security/artikel/98090/0).

Mit 10.5 scheinen sich Apples APIs und Frameworks allmählich zu stabilisieren. Cocoa-Programmierer haben mit minimalem Aufwand Zugriff auf mächtige OS-Funktionen – beispielhaft seien Core-Image und -Animation genannt. Andere Frameworks wie QuartzGL/CoreUI erschließen eine weitere Leopard-Novität, die komplette Auflösungsunabhängigkeit des Aqua-GUI: Control plus Scroll-Rad vergrößert die Ansicht. Über kurz oder lang muss man daher damit rechnen, dass viele Programme 10.5 als Mindestvoraussetzung vorgeben, auch da mit Leopard Signed Code in großem Stil Einzug halten wird, also die Unversehrtheit als Programmcode zur Voraussetzung für den Start gelten muss.

Automator protokolliert Klicks

Aber auch dem Gelegenheitsprogrammierer beziehungsweise Skripter bietet sich viel Neues: Applescript ist in Version 2.0 mächtiger und zugleich komfortabler, zudem lassen sich weitere Bereiche des Systems vollständig damit steuern. Der Automator als von Laien bedienbares Universalwerkzeug für Automatisierung erhielt mehr Nutzwert. So lassen sich mit ihm jetzt Benutzeraktionen wie Mausklicks und Tastendrücke aufzeichnen und kombiniert mit anderen Automator-Aktionen als Workflow speichern.

Python und Ruby können vollwertig für Skripting per Apple Events benutzt werden [2], ab 10.5 geht das auch aus Objective-C-Code heraus – die neue Scripting Bridge macht es möglich. Spätestens nach Installation der kostenlosen Entwicklungsumgebung *XCode 3.0* steht mit Mac OS X 10.5 eine vollwertige und umfangreich ausgestattete Entwicklungsplattform auch für klassische Webanwendungen zur Verfügung. *DTrace* rüstet sie zudem mit maximalem Komfort bei der Analyse aus.

Fazit

Apples OS X 10.5 bringt neben vielen Detailverbesserungen erstmals den kompletten Umstieg auf 64 Bit. An der Oberfläche fällt eine Vereinheitlichung der Fenstergestaltung ins Auge. Das mitgelieferte Backup- und Restore-Programm „Time Maschine“ alleine würde den Umstieg rechtfertigen. Bislang plagen allerdings noch Kinderkrankheiten gerade dieses Produkt. Für den produktiven Einsatz empfiehlt es sich daher, bis zur Version 10.5.3 oder 10.5.4 zu warten. (ck)

THOMAS KAISER

ist selbstständiger Berater/Integrator vor allem für Unix-Server, Macs, Workflow-Optimierung und Druck-/Medienvorstufe.

Literatur

[1] Bert Ungerer; Betriebssysteme/Diverses; Katzennachwuchs; Tiger: Neue Version von Mac OS X; *iX* 6/2005, S. 18

[2] Thomas Kaiser; Scripting; Futter für die Katze; Tiger: Mac OS X mit Unix-Mitteln nutzen; *iX* 2/2006, S. 144

Tiger-Look

Wer als Administrator fürchtet, Anwender wären mit einzelnen Aspekten der neuen Oberfläche überfordert, obwohl die Macht der Gewöhnung eher früher als später Abhilfe schafft, dürfte an Folgendem Gefallen finden:

– Dock von der Verspiegelung befreien:

```
defaults write com.apple.dock no-glass \
-boolean YES; killall Dock
```

(alternativ TinkerTool dafür verwenden, siehe „Onlinequellen“)

– Hierarchische Ordner statt Stacks im Dock: Installation von HierarchicalDock (alternativ DragThing einsetzen oder die Vorzüge der Finder-Seitenleiste schätzen lernen)

– Transluzenz der Menüleiste reduzieren:

```
sudo defaults write ⤦
    /System/Library/LaunchDaemons/ ⤦
    com.apple.WindowServer ⤦
    'EnvironmentVariables' -dict ⤦
    'CI_NO_BACKGROUND_IMAGE' 0.62
```

(alternativ OpaqueMenuBar installieren oder Schreibtisch-Hintergrundbild mit 20 Pixel hohem, hellen Streifen nutzen)

iX-Wertung

⊕ virtuelle Desktops

⊕ auflösungsunabhängiges GUI

⊕ Verfolgung laufender Prozesse per *dtrace*

⊖ Time Machine noch nicht zuverlässig

Lieferumfang und Preise

Betriebssystem	Mac OS X
Version	10.5 „Leopard“
Anbieter	www.apple.de
Medium	DVD
Preis	Einzelbenutzer: 108,40 €; 5 Benutzer (Familienlizenz) 167,23 € zzgl. MwSt.

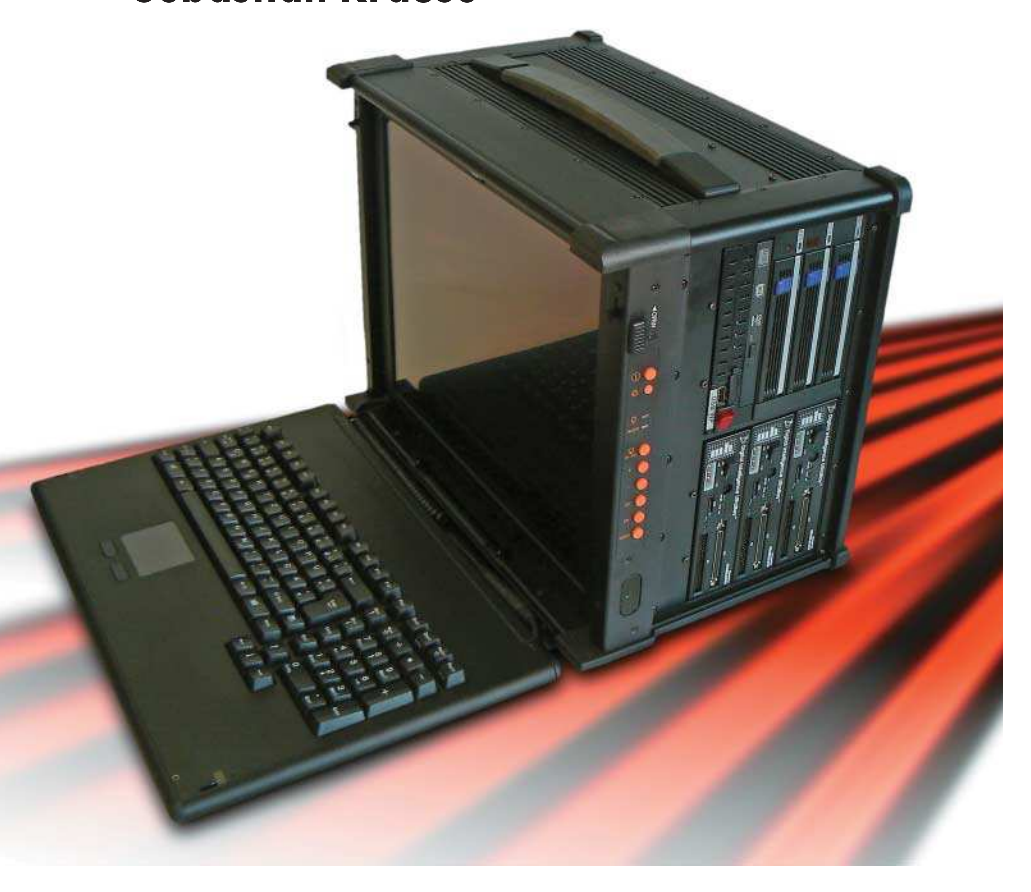
Onlinequellen

Apples Feature-Liste	www.apple.com/de/macosx/features
Ars Technica Review	arstechnica.com/reviews/os/mac-os-x-10-5.ars
TinkerTool	www.bresink.de/osx/TinkerTool-de.html
DragThing	www.dragthing.com
HierarchicalDock und OpaqueMenuBar	eternalstorms.at
Quick Look Plug-Ins	apfelquak.de/ql-plugins

Umfangreiche Datenmengen sichern mit dem TreCorder

Express-Sicherung

Sebastian Krause



Viele Daten in möglichst kurzer Zeit sicherzustellen, ist für Ermittler von Computerdelikten aller Art unerlässlich. Eine neue Hochgeschwindigkeits-Hardware soll das bewerkstelligen und zeigt, was technisch möglich ist.

Die Beweismittelsicherung umfangreicher Datenmengen ist eine Herausforderung für die Behandlung von Sicherheitsvorfällen und die digitale Forensik. Das muss nicht nur schnell, sondern auch fälschungssicher über die Bühne gehen. Ein in Zusammenarbeit von deutschen und Schweizer Ingenieuren entwickelter tragbarer Forensikcomputer namens TreCorder soll insbesondere Sicherheits- und Strafverfolgungsbehörden bei der Erstellung der gerichtsverwertbaren digitalen Kopien in kürzester Zeit unterstützen.

An forensische Duplikate als Basis für die eigentliche Untersuchung, die sogenannte Post-mortem-Analyse, stellen die Beteiligten unter anderem hohe Integritätsanforderungen, nicht zuletzt aus Gründen der späteren Gerichtstauglichkeit. So muss durch Prüfsummenverfahren belegt und jederzeit nachweisbar sein, dass das bitweise erstellte Duplikat eines verdächtigen Datenträgers exakt dem Originaldatenträger entspricht. Schreibzugriffe auf den zu untersuchenden Datenträger soll ein Write-Blocker (Schreibschutz) verhindern.

Lässt sich diese Anforderung bei einzelnen Festplatten aus Arbeitsstationen mit den in Unternehmen und Behörden üblichen Größen noch mit vertretbarem technischen und zeitlichen Aufwand realisieren, wächst der Aufwand bei umfangreichen Speichersystemen wie RAID- oder NAS-Systemen, aber auch groß dimensionierten Festplatten erheblich.

In diesen Fällen spielt der von der deutschen Firma mh SERVICE und der Schweizer Arina Electronic in Zusammenarbeit mit den Landeskriminalämtern und dem Bundeskriminalamt entwickelte TreCorder seine Stärken aus. Der aufgrund des Metallgehäuses sehr robuste, aber auch schwere mobile Computer ist unter anderem ausgestattet mit drei internen Write-Blockern vom Typ T345 des Herstellers Tableau, die den Anschluss von Festplatten nach den Schnittstellenstandards IDE, SATA und SCSI ermöglichen. Die in 5,25“-Schächten eingebauten Write-Blocker sind intern über IEEE 1394b (Firewire 800) angebunden und ermöglichen Hot-Swapping – Ein- und Ausstöpseln im laufenden Betrieb – der zu untersuchenden Datenträger.

Schreibzugriff unterbinden

Nach außen stellen die Write-Blocker jeweils die Stromversorgung für die Festplatten bereit (Abbildung 1). Die für den Anschluss der Festplatten benötigten Kabel werden in einem Alukoffer separat mitgeliefert. In der aktuellen Ausführung sind ebenfalls integrierte schreibgeschützte Schnittstellen für USB und Fibre Channel sowie ein schreibgeschützter 12-in-1-Kartenleser enthalten. Somit ermöglicht es der TreCorder, von allen verbreiteten Speichermedien schreibgeschützt bitweise eine digitale Kopie zu erstellen. Auf kleine Details wie externe Festplattenlüfter haben die Hersteller ebenfalls geachtet. Vor Beschädigungen des TreCorders auf Flügen schützt ein erhältlicher Flight-Case. Jedoch sollte man die Datenträger-Images auf den internen Festplatten des TreCorders verschlüsseln oder zumindest die Wechselfestplatten „am Mann“ transportieren.

Ermittler können den TreCorder sofort einsetzen. Für die Erstellung der forensischen Duplikate sind unter Linux der AIR Imager sowie das EnCase-Werkzeug LinEn und unter Windows der FTK Imager von Accessdata sowie



Die in den Schächten eingebauten Write-Blocker ermöglichen den Anschluss von IDE-, SATA- und SCSI-Festplatten und stellen die Stromversorgung für sie bereit (Abb. 1).



Seine volle Geschwindigkeit erreicht der TreCorder beim parallelen Erstellen von forensischen Duplikaten dreier Festplatten, hier mit dem Werkzeug FTK Imager (Abb. 2).

X-Ways Forensics installiert. Da Letzteres als Vollversion enthalten ist, können die sichergestellten Daten im Anschluss gleich untersucht werden. Aufgrund der Ausstattung mit performanten Systemkomponenten wie dem Quad-Core-Prozessor Q6600 (2,4 GHz) und 4 GByte RAM PC800 eignet sich der TreCorder auch für die Post-mortem-Analyse.

Für die Speicherung der forensischen Images stehen im TreCorder drei Zielfestplatten à 750 GByte im Wechselrahmen zur Verfügung. Diese sind mit dem Dateisystem FAT32 formatiert, um Zugriffe aus Windows und Linux auf die sichergestellten Daten zu ermöglichen. Das bringt die Einschränkung mit sich, dass maximal 4 GByte große Dateien erstellt werden können und die in der Regel deutlich größeren Festplatten-Images geteilt werden müssen.

Eleganter wäre der Einsatz von Ext 2/3, auf das man auch aus Windows mit einem Treiber wie „Ext2 IFS for Windows“ (www.fs-driver.org) lesend und schreibend zugreifen kann. Den TreCorder liefert mh SERVICE als Dual-Boot-System mit Opensuse in der jeweils aktuellen Version (Testsystem mit Version 10.3) und Windows XP Professional SP 2 aus.

So wurde getestet

Der Hersteller gibt für die gleichzeitige Sicherung dreier Festplatten Datenraten von etwa 9,3 GByte pro Minute an. Im Rahmen der Teststellung wurden mit dem TreCorder Datenträger-Images der folgenden Festplattentypen erstellt:
– Samsung SP1614N, IDE-Schnittstelle, 160 GByte, Baujahr 05/2004
– Samsung HD321KJ, SATA-Schnittstelle, 320 GByte, Baujahr 04/2007

Die Messungen erfolgten mit den Werkzeugen AIR Imager und LinEn unter Linux sowie mit FTK Imager unter Windows. Die Geschwindigkeitsvorteile aus der Verwendung der Tableau-Write-Blocker sollte eine Vergleichsmessung über einen schreibgeschützten USB-2.0-Anschluss des TreCorders, das heißt ohne Tableau-T345-Write-Blocker belegen. Des Weiteren kam als alternative Methode ein Dell-Notebook D630 mit der Forensiktool-Sammlung Helix 1.8 und dem Werkzeug LinEn zum Einsatz, um das Image einer über einen Write-Blocker Tableau T5 und IEEE 1394b ange-



Messdaten bei Duplikation verschiedener Festplatten

Messung Nr.	Festplattentyp	Kapazität	Sicherungshardware (Binärpräfixe)	Sicherungssoftware in GByte	Prüfsumme	Dauer (hh:mm:ss)	MByte / Sekunde	GByte / Stunde	GByte / Minute
1	Samsung SP1614N, IDE, 160 GByte	149,05	TreCorder, T345 auf interne SATA-Zielfestplatte	Linux, Linen 5.05c	MD5	00:55:35	45,77	160,89	2,68
2	Samsung SP1614N, IDE, 160 GByte	149,05	TreCorder, T345 auf interne SATA-Zielfestplatte	Linux, AIR-Imager 1.2.8	MD5	01:14:47	34,02	119,59	1,99
3	Samsung SP1614N, IDE, 160 GByte	149,05	TreCorder, T345 auf interne SATA-Zielfestplatte	Windows, FTK Imager 2.5.3	MD5	01:00:01	42,38	149,01	2,48
4	Samsung HD321KJ, SATA, 320 GByte	298,09	TreCorder, T345 auf interne SATA-Zielfestplatte	Linux, Linen 5.05c	MD5	01:40:15	50,75	178,41	2,97
5	Samsung SP1614N, IDE, 160 GByte Samsung SP1614N, IDE, 160 GByte Samsung SP1614N, IDE, 160 GByte	149,05 149,05 149,05	TreCorder, 3xT345 auf interne SATA-Zielfestplatten	Linux, Linen 5.05c	MD5	00:55:20 00:54:40 00:56:42	45,97 46,53 44,86	161,62 163,59 157,73	2,69 2,73 2,63
								Summe:	**8,05**
6	Samsung HD321KJ, SATA, 320 GByte Samsung HD321KJ, SATA, 320 GByte Samsung HD321KJ, SATA, 320 GByte	298,09 298,09 298,09	TreCorder, 3xT345 auf interne SATA-Zielfestplatten	Linux, Linen 5.05c	MD5	01:44:15 01:39:41 01:41:43	48,80 51,04 50,02	171,56 179,42 175,84	2,86 2,99 2,93
								Summe:	**8,78**
7	Samsung SP1614N, IDE, 160 GByte	149,05	TreCorder, Quell-Festplatte USB 2.0 auf interne SATA-Zielfestplatte	Linux, Linen 5.05c	MD5	02:02:58	20,69	72,73	1,21
8	Samsung SP1614N, IDE, 160 GByte	149,05	Dell D630, Quell-Festplatte über Tableau T5 IEEE 1394, Zielfestplatte Samsung SP1614N über ICYBOX (IB-351UE-B USB 2.0)	Helix 1.8 Linux, Linen 5.05f	MD5	02:18:32	18,36	64,56	1,08

schlossenen Festplatte auf eine Zielfestplatte in einem externen USB-2.0-Gehäuse zu erstellen.

Vor jedem Image-Vorgang fand eine Neuformatierung der Zielfestplatten statt, um einheitliche Voraussetzungen zu schaffen. Es wurden bei allen Messungen MD5-Prüfsummen erstellt, auf eine Verifikation jedoch verzichtet. Die Images wurden immer in 2 GByte große Dateien aufgeteilt. Die Ergebnisse der Image-Vorgänge sind in der Tabelle aufgeführt.

Grenzen der Geschwindigkeit

Die Abweichungen der Übertragungsrate zwischen den beiden Festplattentypen bei gleichen Anschlussvarianten verdeutlichen die Abhängigkeit von der Bauart beziehungsweise von Schnittstelle und Baujahr der untersuchenden Festplatte. In der Praxis treffen Ermittler bei vielen Computerdelikten auf ältere Festplatten, die die Geschwindigkeitsreserven des TreCorders nicht ausschöpfen können. Dass die Wahl des Werkzeugs für die Beweismittelsicherung ebenfalls Auswirkungen auf die zu erreichenden Datenraten hat, zeigen die Unterschiede zwischen den verwendeten Hilfsmittel. Das in der etwas älteren Version 5.05c installierte LinEn erreichte hier die besten Ergebnisse (siehe Tabelle).

Der Vergleich der alternativen Methode mit Tableau T5 am Dell-Notebook und dem Anschluss über den Write-Blocker T345 des TreCorders offenbart einen fast 2,5-fachen Geschwindigkeitszuwachs bei Einsatz des TreCorders. Wenn auch am TreCorder nur der schreibgeschützte USB-2.0-Anschluss genutzt wird, verringert sich der Vorsprung auf etwa 2 MByte pro Sekunde. Der TreCorder erreichte erwartungsgemäß seine maximale Effektivität bei der parallelen Beweismittelsicherung von drei Festplatten gleichzeitig. Hier wurde ein Spitzenwert von 8,78 GByte pro Minute gemessen. Weitere Messergebnisse aus herstellereigenen Tests in Zusammenarbeit mit dem LKA Niedersachsen sind auf der Webseite der Firma mh SERVICE (www.mh-service.de) veröffentlicht.

Fazit

Insbesondere für Sicherheits- und Strafverfolgungsbehörden, die sich regelmäßig und unter Zeitdruck mit umfangreichen Datenmengen im Rahmen von computerforensischen Untersuchungen konfrontiert sehen, beschleunigt der TreCorder die gerichtsverwertbare Beweismittelsicherung deutlich. Gerade wenn mehrere Festplatten sichergestellt werden müssen, kann er seine Stärken durch die parallele Imageerstellung voll ausspielen. Für Ermittler, die etwas mehr Zeit mitbringen, gibt es allerdings auch deutlich günstigere Alternativen wie das Anschließen von externen Write-Blockern über Firewire 800 oder USB 2.0 an performante Notebooks. (ur)

SEBASTIAN KRAUSE

ist Security Consultant bei der Berliner HiSolutions AG.

Literatur

[1] Sebastian Krause; Computerforensik; Im Notfall; Sofortmaßnahmen nach einem Systemeinbruch; *iX* 7/2007, S. 90

[2] Sebastian Krause; Computerkriminalität; Verdeckte Ermittlung; Shadow 2: Forensik-Hardware zur Systemanalyse; *iX* 10/2006, S. 72

iX-Wertung

⊕ schnelle Beweissicherung umfangreicher Datenmengen

⊕ robuste Verarbeitung

⊕ Hard- und Software für sofortigen Einsatz vorhanden und installiert

⊖ schwer und unhandlich

Daten und Preise

TreCorder,
Forensik-Hardware

Hersteller: mh SERVICE GmbH

Website: www.mh-service.de

Preis: 9800 Euro

Lüfterloser Tischrechner Senyo 710 von Transtec

Im Stillen

Ralph Hülsenbusch



Transtec hat einen Desktop-Rechner auf den Markt gebracht, der nicht nur geräuschlos sein, sondern zudem mit der Energie haushalten und trotzdem die für einen PC akzeptable Leistung erbringen soll.

Ein wenig Retrolook, der manch einen an die Kühlkörper der mit Transistoren bestückten Hi-Fi-Endstufen erinnern dürfte, das fällt beim Senyo 710 als erster Unterschied zu bisherigen Desktop-Rechnern ins Auge. Andere sehen vielleicht Parallelen zum alten Radio – nein, die beiden Knöpfe kann man nicht drehen. Der rechte Taster dient zum Ein-/Ausschalten, der linke zum Auswerfen der Laserdisk. Die leuchtenden Ringe signalisieren rechts eisblau „eingeschaltet“ und links flackernde Aktivitäten im Rotlicht.

Dass es sich bei dem Gehäuse im Grunde genommen um einen einzigen Kühlkörper handelt, verrät einem der Warnhinweis auf dem Aufkleber: Man darf da nichts draufstellen – ein erstes Zugeständnis an den Lüfterlosen. Grundsätzlich sollte man solchen Geräten, die ihre innere Hitze über die Außenhaut abstrahlen müssen, ringsrum Luft lassen.

Besondere Umgebungen

Transtec sieht die Haupteinsatzbereiche in Büros und daheim als Mediacenter sowie in Bereichen, in denen es auf Ruhe ankommt, etwa in Arztpraxen und Krankenhäusern. Für die übliche Büroarbeit und das Surfen im Netz ist selbst die kleinere Konfiguration mehr als hinreichend ausgestattet, wie in der Tabelle rechts zu sehen.

Hervorzuheben sind der Netzschalter, der das Gerät vom Stromnetz trennt, die reichliche Zahl von Anschlüssen, vor allem der für DVI, zu bemängeln das magere Angebot für USB-Geräte. Um- oder Ausbaumaßnahmen erschwert Transtec durch spezielle Sternschrauben, für die nur wenige den passenden Schraubendreher haben dürften. Dafür verhindert das rundum geschlossene Gehäuse das Eindringen von Staub. Gespart hat der Hersteller bei den transparenten Kunststoffnoppen unter dem Boden, die ein Hin- und Herrutschen verhindern. Im Test war nach kurzer Zeit einer abgefallen und verschwunden.

Für den Einsatz als Multimediacenter setzt die Grafik Grenzen. Intels GMA 950 eignet sich wegen des „dynamischen“ Speichers, den der Controller vom Hauptspeicher abzwackt, nicht für anspruchsvolle 3D-Darstellungen. Messungen mit einigen Benchmarks zeigen, dass der Senyo mit der Core 2 Duo CPU T7200 nicht viel mehr Leistung als mit dem Celeron bringt. Das sonst nahezu geräuschlose Gerät beginnt heftig zu lärmen, wenn es für eine CD das Laufwerk auf Touren bringt. Und nicht nur das: Es riecht unangenehm nach überhitzter Elektronik. Im längeren Betrieb steigt die Temperatur des Gehäuses auf über 40° C (im Test auf 44° C, an der Rückseite gemessen).

Im Standby ergaben Messungen einen Verbrauch von 13,5 W, unter Volllast stieg der Wert auf knapp 47 W. Damit liegt der Senyo unter den Werten handelsüblicher Notebooks. Aber auf der Kehrseite der Medaille steht ein Wirkungsgrad von etwas über 50 %. Das heißt, die Versorger müssen fast doppelt so viel liefern wie das Gerät an Wirkleistung bietet. Beim Hochfahren kommen Spitzen von bis zu 72 W zustande, vollständig ausgeschaltet bleibt ein Rest von 2 VA Blindleistung.

Lieferumfang, Preise und *iX*-Wertung

Senyo 710 PC, lüfterloser Desktop-Rechner

Hardware: Core 2 Duo T7200, 2 GHz, 2 MByte L2-Cache; 2 GByte Dual-Channel DDR2 SDRAM; 160 GByte ATA-HD; DVD-Laufwerk: Double-Layer, +/–RW; GMA950 GM onboard Grafik; 3 × Gigabit-Ethernet; Anschlüsse Rückseite: 2 × PS/2, 3 × RJ45 (Ethernet), 6 × Audio, je 1 × RS-232, IEEE 1394 (Firewire), VGA, DVI und Strom mit Ein-/Ausschalter

Software: Windows XP Professional 2002 SP2; Handbuch (PDF), Nero Express 7

Hersteller/Anbieter: Transtec, Tübingen (www.transtec.de)

Preis: 2260 € (Teststellung)

***iX*-Wertung:**
⊕ geräuschlos (ohne CD) ⊖ schwache Grafikleistung ⊖ nicht erweiterbar
⊕ Strom sparend ⊖ Emissionen bei DVD-Nutzung ⊖ zu teuer

Fazit

Fanless stellt sicherlich ein besonderes Merkmal für Desktop-PCs dar. Der eine oder andere dürfte sich für den geringen Stromverbrauch und fast völlige Stille beim Arbeiten begeistern. Die Technik ist allerdings im industriellen und klinischen Bereich bekannt und kommt teilweise in ruggedized Notebooks zum Einsatz [1]. Und auch dort gilt, dass die Geräte im Vergleich zu herkömmlichen Desktops mit geringerer Leistung arbeiten. Fans von Musik und Videos auf DVD oder CD kann Transtec mit dem Senyo nicht gewinnen. Die Geräusche übertönen bei CDs sanfte Passagen und ob die Geruchsbelästigung sich mit der Zeit verliert, bleibt fraglich. Noch offen bleibt die Frage, warum der Senyo fast doppelt so teuer ist wie die vergleichbaren Mac oder PC Minis [2]. Alles in allem ist aber der Trend zum sparsamen Umgang mit Energie zu begrüßen und die damit verbundene Rückkehr des Ausschalters. (rh)

Literatur

[1] Peter Köhler; Robuste Rechner; Normlage; Normen und Testbedingungen für ruggedized Notebooks; *iX* 12/2007, S. 92

[2] Ralph Hülsenbusch; Systeme; PC-Anlauf gegen Mac Mini; *iX* 1/2006, S. 42



Passwörter knacken mit Elcomsofts DPR Suite

Der letzte Zugriff

Alexander Geschonneck

Für jede Software gibt es einen „good use“ und einen „bad use“. Das gilt besonders für heikle Programme wie Passwortknacker – doch ob Spitzbube oder Administrator, beide wollen, dass er funktioniert.

Der Zugriff auf passwortgeschützte Daten ohne Kenntnis des dazugehörigen Kennwortes kann mitunter auch ohne böse Absichten nötig sein. Es gibt viele legale Anwendungszwecke für Passwortanalyseprogramme. Wurde ein Passwort vergessen oder ist der einzige Administrator, der es kannte, nicht mehr im Unternehmen, muss man an die verschlüsselten Daten mit Gewalt herankommen. Das betrifft auch den Bereich der computerforensischen Ermittlung, wenn sichergestellte Datenträger oder darauf gespeicherte Anwendungen und Daten mit Passwörtern geschützt sind. Der russische Softwarehersteller Elcomsoft hat seine Passwortanalyseprogramme in einer neuen Sammlung zusammengefasst. Die im Elcomsoft Passwort Recovery Bundle für Ermittler in Version 2.0 enthaltene Distributed Passwort Recovery Suite (DPR) bietet nun auch die Möglichkeit, in verteilten Umgebungen komplexe Passwortanalyseverfahren durchzuführen. Eine weitere Neuerung ist die Option, Windows-SYSKEY-Startup-Passwörter sowie zwischengespeicherte Domain Credentials zu analysieren. Und schließlich kann die neue Software aufwendige Passwortanalyseaufgaben durch eine GeForce-8-Grafikarte erledigen lassen – ein Verfahren, das das Unternehmen zum Patent angemeldet hat.

DPR besteht im Wesentlichen aus drei Komponenten: Agent, Server und Konsole. Sie lassen sich unabhängig voneinander auf verschiedenen Systemen installieren und betreiben und damit sowohl im LAN als auch in größeren Netzwerken einsetzen. Die Authentifizierung und Kommunikation zwischen allen DPR-Komponenten erfolgt verschlüsselt. Ob das verwendete Verschlüsselungsverfahren durch die DPR selbst knackbar ist, geht aus der Dokumentation nicht hervor. Da letztendlich vertrauliche Informationen wiederhergestellt werden, ist es sowieso ratsam, in einer geschützten Umgebung zu arbeiten.

Ist die DPR-Serverkomponente gestartet, lässt sie sich von jedem anderen System im Netz mit der darauf befindlichen Konsole administrieren. Über die Konsole legt der Benutzer die Passwortanalysevorgänge auf dem Server fest und startet sie. Der Server verteilt diese sogenannten Tasks dann an die im Netz befindlichen DPR-Agents. Die Aufgabenverteilung ermöglicht ein unabhängiges Arbeiten der Agents – was sich unter anderem günstig auf die Performance auswirkt.

Die Agenten führen ihre Analysetasks in der Idle-Time des Systems aus, auf dem sie installiert sind. Bricht die Verbindung ab oder startet der Server neu, arbeiten sie trotzdem weiter und liefern ihre Analyseergebnisse ab, wenn der Server wieder erreichbar ist. Wenn der Server erreichbar ist, liefern die Agents

Unterstützte Anwendungen und Dateiformate

- Microsoft Word/Excel/PowerPoint 2007 (.DOCX, .XLSX, .PPTX) (Passwortwiederherstellung – nur das „öffnen“-Passwort)
- Microsoft Word/Excel/PowerPoint XP/2003 (.DOC, .XLS, .PPT) (Passwortwiederherstellung – nur das „öffnen“-Passwort)
- Microsoft Word/Excel 97/2000 (.DOC, .XLS) (Passwortwiederherstellung – nur das „öffnen“-Passwort)
- Microsoft Word/Excel 97/2000 (.DOC, .XLS) (Entschlüsselung)
- Microsoft Money (Passwortwiederherstellung)
- Microsoft OneNote (Passwortwiederherstellung)
- PGP ZIP Archive (.PGP) (Passwortwiederherstellung)
- PGP Secret Key Rings (.SKR) (Passphrasewiederherstellung)
- PGP Disks mit konventioneller Verschlüsselung (.PGD) (Passwortwiederherstellung)
- PGP Selbstentschlüsselnde Archive (.EXE) (Passwortwiederherstellung)
- PGP Whole Disk Encryption (Passwortwiederherstellung)
- Personal Information Exchange Zertifikate – PKCS#12 (.PFX, .P12) (Passwortwiederherstellung)
- Adobe Acrobat PDF mit 128-Bit-Verschlüsselung (Passwortwiederherstellung)
- Adobe Acrobat PDF mit 40-Bit-Verschlüsselung (Passwortwiederherstellung)
- Adobe Acrobat PDF mit 40-Bit-Verschlüsselung (Entschlüsselung)
- Windows NT/2000/XP/2003/Vista logon passwords (LM/NTLM) (Passwortwiederherstellung)
- Windows SYSKEY Startup Passwort (Passwortwiederherstellung)
- Windows DCC (Domain Cached Credentials) Passwörter (Passwortwiederherstellung)
- Intuit Quicken (.QDF) (Passwortwiederherstellung)
- Lotus Notes ID Dateien (Passwortwiederherstellung)
- MD5 Hashes (Klartextwiederherstellung)

Angaben laut Hersteller

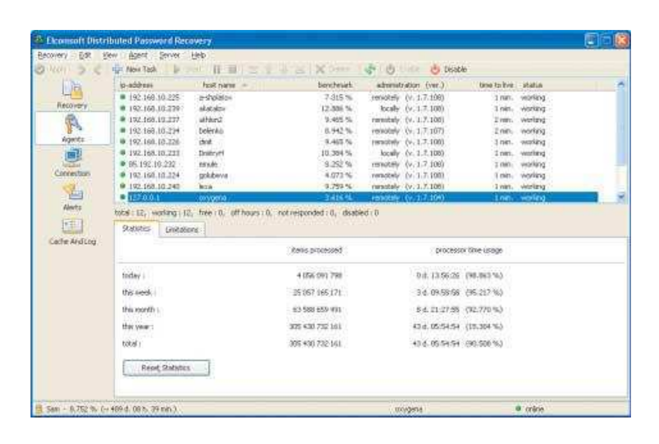
minütlich Statusinformationen, die man über die Konsole auslesen kann. Der Server kann allerdings nur solchen Agents Analyseaufgaben zuweisen, die sich beim Start bei ihm registriert haben.

Verteilte Kraft

Halten sich Anwender an Security-Policies, können Passwörter lang und komplex sein. Ein einfacher Brute-Force-Angriff eines einzelnen Computers reicht dann nicht aus, sie in akzeptabler Zeit zu brechen. Das Durchprobieren aller möglichen Zeichen- und Buchstabenkombinationen dauert einfach zu lange. Durch das Verteilen der Passwortanalyse auf viele Rechner lässt sich wertvolle Zeit gewinnen.

Via DPR-Konsole lassen sich weitere DPR-Agents einfach installieren oder deinstallieren und auch das Rechenverhalten sowie die zu nutzenden Rechenressourcen und -prioritäten einstellen. So kann man etwa Wochentag und Uhrzeit festlegen, zu der Passwortanalysetätigkeiten durch den Agent auf dem Client gestartet werden dürfen. Ebenfalls zentral für jeden einzelnen Client einstellbar ist, ob der DPR-Agent als Systemdienst starten soll. Zu den laufenden Passwortanalysetasks lassen sich über die zentrale Konsole Protokollinformationen und Statistiken abrufen.

Zur diesjährigen Systems hat der Hersteller eine neue Plattform für die Passwortanalyse gezeigt. Es ist mit der DPR-Version 2.0 nun möglich, rechenaufwendige Tasks durch eine GeForce-Grafikkarte des Herstellers Nvidia erledigen zu lassen. Laut Hersteller lassen sich beim Analysieren von NTLM-Hashes Geschwindigkeitsvorteile von bis zu 25 Prozent erzielen – bei Brute-Force-Analysen ein spürbarer Gewinn. GeForce-8-Karten werden sich im Büroumfeld aber wohl eher selten finden lassen.



Die zentrale Konsole ermöglicht den einfachen und schnellen Zugriff auf alle Komponenten und bietet einen Überblick über die laufenden Tasks (Abb. 1).

Fazit

Passwortanalysen sind nicht nur für Bösewichte sinnvoll. Hat ein Mitarbeiter oder Administrator sein Passwort vergessen, ist Passwortknacken zugegebenermaßen nicht die beste, manchmal aber die einzige Notfallmaßnahme. Auch im Bereich der computerforensischen Auswertung von verschlüsselten Dateien stellt sich oft die Frage der Zugriffsmöglichkeiten. Die Distributed Password Recovery Suite von ElcomSoft bietet die Möglichkeit, auch komplexe und langwierige Passwortanalysen in einer verteilten Umgebung zu beschleunigen. (ur)

ALEXANDER GESCHONNECK

ist leitender Sicherheitsberater bei der HiSolutions AG in Berlin und Autor des Buches „Computer-Forensik – Systemeinbrüche erkennen, ermitteln, aufklären“.

Daten und Preise

Distributed Password Recovery Suite

Hersteller: ElcomSoft Co Ltd, www.elcomsoft.com

Betriebssystem: Windows 2000, XP, Vista, 2003 Server

Preis: je nach Anzahl der von der Konsole zu bedienenden Agents, bis zu 100 Clients 1199 Euro

Sprachen: Deutsch, Englisch, Russisch, Japanisch

iX-Wertung

⊕ verteilte Analyse spart Zeit

⊕ ressourcenschonender Grafikkarteneinsatz

⊕ unterstützt sehr viele Formate und Anwendungen

⊖ für Agents nur Windows-Versionen verfügbar





Hochgerechnet

Ralph Hülsenbusch

Zunehmend rückt der Aspekt des Energieverbrauchs in der IT in den Vordergrund, selbst bei Rechnern der oberen Leistungsklasse. Worauf es dabei ankommt, will Lynx mit seinem Server und neuen Energiespar-CPUs zeigen.

Wenn heutzutage von Multiprozessor-Rechnern die Rede ist, kommt es leicht zu Missverständnissen, denn in jedem Prozessorchip werkeln mehrere Kerne, die oft als CPU bezeichnet werden. Intels jüngste Xeon-Generation besitzt vier davon und kommt in Systemen mit mehreren Sockeln zum Einsatz – bis auf die U-Version für Single-Boards [1]. Der Server, den Lynx zum Test lieferte, besitzt zwei Sockel und hat damit acht CPU-Kerne.

Das erfordert eine entsprechende Ausrüstung mit Speicher, denn angesichts heutiger Datenvolumina brauchen Anwendungen pro Kern gut und gerne 1 GByte Hauptspeicher, manche sogar noch etwas mehr. Und für Server-Applikationen gilt beim Speicher: je größer, desto schneller. Die muss wiederum ein genügend großer und gesicherter Massenspeicher versorgen. Die Größe der Festplatten spielt dabei weniger eine Rolle als vielmehr deren Zahl, denn nur mit Redundanz sind RAID-Level realisierbar, die den Ansprüchen an Verfügbarkeit der Dienste gerecht werden können.

Energiesparen mit speziellen CPUs

Lynx lieferte seinen Workgroup Server LYSRV-5I2RAB mit zwei Xeon Quad-Core E5345 (Clovertown) und 16 GByte Hauptspeicher sowie Suse Linux Enterprise Server 10 (SLES 10) in der 64-Bit-Ausführung. Die 80-Watt-Prozessoren sind mit 2,3 GHz getaktet, verfügen über zwei 4 MByte große L2-Caches und einen 1333 MHz schnellen Frontside Bus (FSB). Im Grunde genommen handelt es sich um zwei Dual-Core-Prozessor-Dies mit je einem L2-Cache in einem Gehäuse – ein Kunstgriff, den Intel schon einmal beim Core 2 Quad verwendet hat. In den Prozessorsockel LGA 771 (Land Grid Array) passen auch andere Xeon CPUs, von denen es unter anderem eine Energiesparversion (Low Voltage) gibt. Lynx hatte ein Pärchen L5335 geliefert, die nur 50 W verbrauchen (siehe Abb. 3).

Der Hersteller, der im Oktober 2006 in den Besitz der Triple Stor GmbH in Tübingen übergegangen ist, verwendet für den Rack-Server Intels fünf Einheiten hohen Barebone 5400LX mit zwei 830-Watt-Netzteilen. Er ist für das Board S5000PSLSAS von Intel ausgelegt, das mit dem 5000P-Chipsatz („Blackford“) bestückt ist. Die Bensley-Plattform ist seit 2006 auf dem Markt und kam zuerst für die neuen Dual-Core-Prozessoren „Dempsey“ heraus. Die folgende Generation „Woodcrest“ nutzt ebenfalls diese Architektur. Es geht darum, den Engpass FSB zu beseitigen, denn mit der Zunahme der Prozessorkerne kommt es zu konkurrierenden Zugriffen auf den FSB und damit zu Wartezyklen. Dasselbe gilt für die Verbindungen zum Hauptspeicher. Je mehr Aufgaben gleichzeitig zu berechnen sind, desto häufiger muss der Chipset dem Prozessor mit seinen Rechenkernen neue Daten liefern, und der muss warten. Untätige Schaltkreise verursachen Kosten, gehen zu Lasten der Performance und verbrauchen Energie.

Beim Blackford-Chipsatz lenkt Intel den Datenverkehr über mehrere Busse. Jeder Prozessor-Chip hat seinen eigenen FSB, der zudem mit einer Frequenz von 1066 oder 1333 MHz über der des Vorgängers „Lindenhurst“ mit 800 MHz liegt. Rein rechnerisch steigt der Datendurchsatz damit von 6,4 GByte/s auf 8,5 respektive 10,7 GByte/s pro FSB. Die Steigerung trifft aber auf die vier Cores, die sich den FSB teilen müssen. Der Zuwachs gegenüber klassischen Vierprozessorsystemen (mit Single Cores) würde schlicht verpuffen, wenn Intel nicht zugleich die Zahl der Speicherbusse verdoppelt hätte: beim 5000P Chipsatz vier anstatt zwei wie beim 5000X. Hinzu kommt, dass hier Fully Buffered Dual Inline Memory Modules (FB DIMM) zum Einsatz kommen, die dank der seriellen Ansteuerung über 24 Leitungspaare eine bis zu viermal schnellere Datenübertragung als bei den Vorgängern, den registered Moduls erlaubt. Außerdem kann ein Bus bis zu acht Speicherbänke ankoppeln. Allerdings braucht die FB-Technik mehr Energie und Kühlung.

Ungebrochener Trend zur Größe

Wer sich auf Zeiten zurückbesinnt, zu denen Digital Semiconductor Alpha-

CPUs (21264) samt Chipsets produzierte, dem dürfte auffallen, dass DEC 1998 mit Tsunami und 1999 mit Typhoon ein ganz ähnliches Konzept mit mehreren Frontside- und Speicherbussen realisiert hatte. IBM hat ein ähnliches Konzept bereits bei seinem 2005 auf den Markt gebrachten X3-Chipset (Hurrican) eingeführt.

Mit der Masse an Prozessorkernen, Speicherbausteinen und Festplatten wächst der Energieverbrauch. Denkt man an ein Acht-Wege-System mit 16 GByte RAM und sechs SAS-Festplatten klassischer Bauart, rechnet man mit einem Verbrauch in der Kilowatt-Klasse. Um dem Energiebedarf des Servers von Lynx auf die Schliche zu kommen, kam im *iX*-Labor ein geprüftes Einphasenmessgerät aus dem Fundus von c't zum Einsatz. Der Messvorgang durchlief vier Phasen: am ausgeschalteten Server, beim Hochfahren, im „normalen“ Systembetrieb ohne Zugriffe von außen und unter Last, erzeugt mit SPECs CPU2006, da es ja vor allem um den Einfluss der Quad-Cores auf den Energieverbrauch geht.

Im „kalten“ Zustand tritt der bekannte unerfreuliche Effekt auf, dass der Server bei einem Verbrauch von 26 W tatsächlich 123 VA Scheinleistung abfordert (Cosinus Phi 0,22, Anteil der Wirkleistung). Das gibt sich jedoch, sobald der Server seine Dienste aufnimmt. Beim Hochfahren kommt es zu Spitzen von über 320 W, der Anteil liegt dabei aber zwischen 0,89 und 0,92. Unbeansprucht nimmt der Server rund 300 W auf und pendelt sich bei 0,92 ein, der Anteil von Scheinleistung bleibt unter Volllast bei 380 W in etwa gleich. Für die Nutzung von acht Prozessorkernen mit 16 GByte Speicher den Strom bezahlen zu müssen, wie ihn vier Glühlampen verbrauchen, klingt durchaus akzeptabel.

Nach dem Prozessorumbau des Servers von Lynx auf zwei Dual-Core-Xeon L5335 (2 GHz) wiesen die Messungen des Energieverbrauchs auf eine leichte Verbesserung hin: mit 270 W (0,87) 10 % weniger und unter Last 330 W (0,90) eine Ersparnis von rund 13 %. Das Runterschalten auf je einen aktiven Kern pro CPU – quasi ein Solo-Prozessorbetrieb – brachte keinerlei maßgebliche Veränderung mit sich. Setzt man aber die Rechenleistung laut CPU2006 ins Verhältnis zum Energieverbrauch, schneidet die LV-CPU etwas besser ab. Das Taktverhältnis 2,3 zu 2 GHz ergibt einen Faktor von 0,86, die Leistungseinbußen liegen aber um die 10 %, während der Energieverbrauch wie der CPU-Takt um 14 % abnimmt. Unter dem Strich ist die Low-Voltage-CPU effektiver.

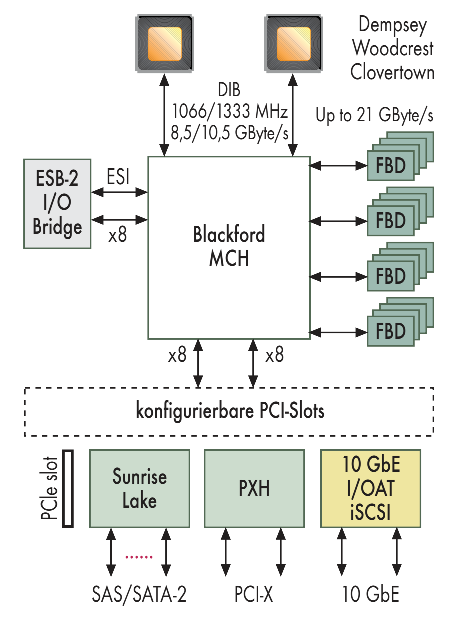


Windkanal: Die Lüfter müssen nicht nur die CPUs, sondern auch die Speicherbausteine kühlen, denn FB-DIMMs strahlen einiges an Hitze ab. Der Luftschacht wurde für das Foto entfernt (Abb. 1).

Positive Bilanz nur bei Low Voltage

Völlig aus dem Rahmen fällt Lynx' Server mit seinem Energieverbrauch nicht. Ein zeitgleich im Test befindlicher 1U-Rechner von ICO mit zwei Quad-Core-Opteron (Typ 2347, 1,9 GHz) und ebenfalls 16 GByte RAM kam in den vier Disziplinen auf ähnliche Werte: 21 W (0,30) ausgeschaltet, 280 W (0,90) im lastfreien Betrieb und 360 W (0,91) bei Beanspruchung. Ein kleineres Modell von Fujitsu-Siemens mit zwei Dual-Core Opteron (Typ 2220, 2,8 GHz) und 8 GByte RAM verbraucht mit schlafenden Prozessoren 195 W und überstieg unter Last kaum die 230-W-Marke.

Offensichtlich sind bei der Wahl der CPU keine allzu großen Einsparungspotenziale gegeben, die liegen eher auf der Softwareseite. Multicore-Systeme sind wegen der Zahl der Prozessorkerne und der Größe des Hauptspeichers prädestiniert als Plattform für virtuelle Maschinen. Hinzu kommt, dass sich Hochverfügbarkeit per Redundanz nach dem N+1-Prinzip herstellen lässt: Ein virtueller Reserve-Server reicht, um einen der ins Trudeln gekommenen zu übernehmen. Zusätzliche Sicherheit bietet die Doppelung der Hardware, bei der die Backup-Server auf der Nachbarhardware installiert sind.



DIB: Dual Independent Bus
MCH: Memory Controller Hub
GbE: Gigabit-Ethernet
FBD: Fully Buffered Dual Inline Memory Modul (FB-DIMM)
ESB: Embedded System Board

Datenwege: Durch den zweifachen Frontside-Bus und den vierfachen Memory-Bus versucht Intel die bisherigen Bottlenecks zu umgehen (Abb. 2).

iX-TRACT

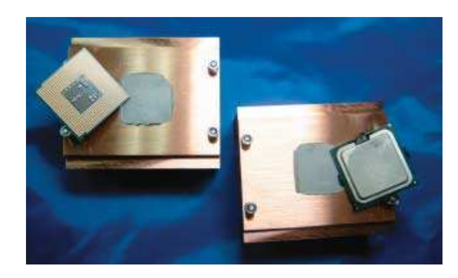
- Multi-Core-Prozessoren helfen Energie in der Produktion zu sparen, da im Unterschied zu früheren Multiprozessoren Boards, Lüfter und zusätzliche Chips nicht mehrfach vorhanden sein müssen.
- Da sie aber mehr und vor allem schnelleren Speicher brauchen, gibt es nur geringe Einsparungspotenziale im Energieverbrauch.
- Bei den Quad-Core-CPUs umgeht Intel Engpässe im Datenaustausch zwischen Rechenkernen und Speicher durch eine Vervielfältigung der Datenbusse.



Handwerk: Zum Test lieferte Lynx zwei Xeons in der Energiesparversion vom Typ L5335, die mit 40 W auskommen (Abb. 3).

Bei der Rechenleistung liefern die Maschinen Werte, mit denen sich mehrere virtuelle Maschinen (VM) betreiben lassen. Zwar kommt keine lineare Skalierung zustande, das heißt acht Core liefern nur etwa die vier- (beim Gleitkommarechnen) bis fünffache Leistung – Ein Manko, mit dem Intels SMP-Systeme immer wieder zu kämpfen haben. Offensichtlich hat AMD im Bereich der Dual-Cores bisher die Nase vorn und kommt mit acht Kernen auf die rund siebenfache Leistung, wobei noch nicht zu erkennen ist, ob das auch für deren Quad-Core-Opterons (Barcelona) gelten wird. Vergleichszahlen von IBM und Supermicro sind bei der SPEC als „not compliant“ (NC), als nicht gültig, markiert. Die RISC-Konkurrenten Power6 von IBM oder Suns Sparc skalieren deutlich besser, wie man in den Ergebnissen der CPU2006 unschwer nachlesen kann.

Was aber ins Auge springt, sind die Unterschiede zu den von Supermicro bei der SPEC publizierten Ergebnissen. Ein Grund liegt darin, dass die Hersteller bei der SPEC gemischte 32- und 64-Bit-Tests und einen neueren Kernel (2.16.43 statt 2.16.21) verwendet haben. Den entscheidenden Unterschied erkennt man aber erst beim Blick auf



Verkittet: Die Techniker haben beim Zusammenbauen die Wärmeleitpaste wie Kitt verwendet. Solche Mengen sollte man tunlichst vermeiden (Abb. 4).

die Boards: Supermicro hat seinen X7DB8+ mit 16 Speicherbänken ausgestattet, während das Original-Intel-Board nur 8 nutzt.

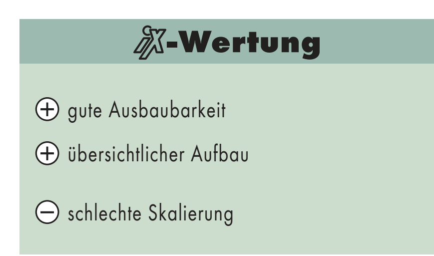
An der Verarbeitung gibt es so gut wie keine Beanstandungen. Der Deckel des Rechners lässt sich nach Lösen einer Verriegelungsschraube leicht vom Gehäuse schieben. Sämtliche Baugruppen sind gut zugänglich und eindeutig gekennzeichnet. Einzig die Verlegung des ATA-Kabels zum DVD-Laufwerk wirkt etwas schlampig und beim Aufbringen der Wärmeleitpaste auf die Prozessoren haben die Techniker bei Lynx arg auf die Tube gedrückt (siehe Abb. 4).

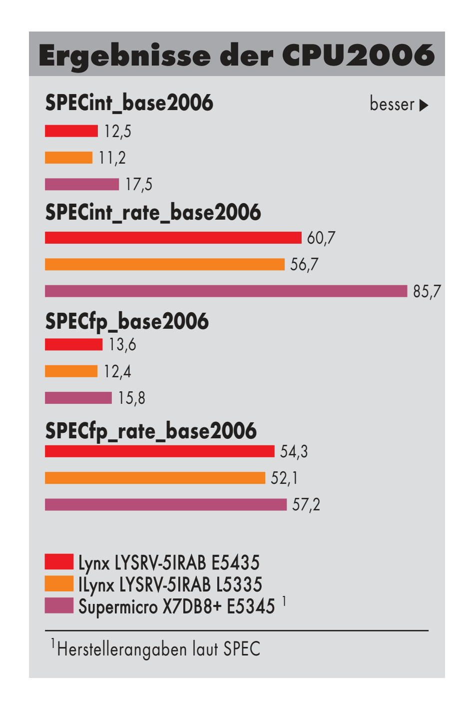
Im Rechner bietet ein integrierter Serviceprozessor den Zugriff über ein separates Netz. Er stützt sich auf IPMI 2.0 und erlaubt dem Administrator über Intels Remote Management Modul (iRMM) per Webserver den Zugriff auf das System. iRMM liefert eine Reihe physikalischer Daten, stellt Java-getrieben eine Konsole und virtuelle Medien (DVD/CD sowie Diskette) bereit, sodass der Admin seine Arbeiten zum großen Teil aus der Ferne erledigen kann. Außerdem kann er den Rechner ferngesteuert hoch- und runterfahren.

Fazit

Für den Einsatz der neuen Quad-Core-Prozessoren von Intel spricht die Zahl der verfügbaren Rechenkerne, nicht aber – oder noch nicht – die Vervielfachung der Rechenleistung. Virtualisierte Umgebungen profitieren von der Zahl der Cores und der Größe des Hauptspeicher.

Bei der Energiebilanz ist es nicht die Hardware, die den Haupteinfluss hat, vielmehr entscheidet, wie der Rechner im tagtäglichen Gebrauch ausgelastet ist. Der Unterschied im Energieverbrauch zwischen Leerlauf und Höchstbelastung liegt bei fast 25 %. Bei einem Konzept, das eine Konsolidierung von Servern mit unterschiedlichen Aufgaben in Form einer Virtualisierung verfolgt und einen hohen Nutzungsgrad der Systeme erreichen will, stehen dem Mehrverbrauch die Einsparung kompletter Rechner gegenüber. Das Plus bei den neuen Quad-Core-Servern steht vor dem großen Hauptspeicher und bei einem 5U-Racksystem den per RAID gesicherten Plattenvolumina. (rh)





Literatur

[1] Ralph Hülsenbusch; Green Computing; Auf Sparkurs; Fujitsu Siemens Computers’ grüner Primergy-Server TX120; *iX* 12/2007, S. 70

[2] Susanne Nolte; Prozessoren; Kerngeschäft; Quad-Core-Rechenwerke für Vier-Sockel-Server; *iX* 11/2007, S. 34

Lieferumfang und Preise

LYSRV-5U2RAB: Quad-Core-Dual-Server

Hardware: zwei E5345 (2,3 GHz), alternativ L5335 (2 GHz) Quad-Core-Xeon mit 2 × 4 MByte L2-Cache, 1333 MHz FSB, 5000p-Chipsatz, acht 2 GByte FD-DIMM, ECC, CL5 667 MHz; drei Fujitsu MAX, SAS, 73 GByte, 16 MByte Cache, 15 000 U/min; Intel, SRCSAS144E, SAS/SATA, PCIe, 4/4 Kanal, RAID 0,1,5,10,50, 256 MByte Cache; DVD-ROM, 16x/48x, ATAPI; zwei Netzteile (redundant) je 830 W

Software: Suse Linux Enterprise Server 10 OEM (1 Jahr Upgrade Protection), DVD-Player PowerDVD 7; Lynx CD-Booklet

Service: Express Exchange, zwei Jahre Austauschdienst

Hersteller: Lynx, Triple Stor GmbH in Tübingen, www.lynx-pc.de

Preis: ab 2982 €

NAS-Cluster mit Infiniband von Isilon



Plattenschwarm

Fred Hantelmann

Vor allem beim High Performance Computing ist der Bedarf nach hohem I/O-Durchsatz groß. Dem begegnen die Hersteller mit dem Konzept des Clustered Storage, das neben der Kapazität auch die I/O-Last auf mehrere gleichberechtigte Knoten verteilt. Ein Neuling auf dem deutschen Markt ist der NAS-Hersteller Isilon.

Zentralisierung von Storage und damit Entkopplung des Massenspeichers von der übrigen Systemhardware locken die geneigte Kundschaft mit dem Versprechen, Ressourcen optimal zu nutzen und gewaltige Einsparungen im Betrieb zu ermöglichen. Gerne zeigen Hersteller und Berater dazu auf blockbasierte Storage-Virtualisierung, die SAN-Lösungen (Storage Area Network) implementieren. Wer die damit verbundene Investition für eine zusätzliche SAN-Infrastruktur scheut, dem bietet dateibasierte Speicher-Virtualisierung per Network Attached Storage (NAS) eine kostengünstige Alternative. Konzeptionelle Mehrwerte von SAN-Produkten gegenüber NAS, etwa Migration, Replikation, Snapshots und effizientes Disaster Recovery, zählen mittlerweile auch bei manchem NAS-Hersteller zum Leistungsumfang.

Sogar in Sachen Durchsatz sieht sich die NAS-Fraktion gegenüber der SAN-Welt inzwischen als gleichwertig oder gar überlegen. Mit dem Argument des minimalen Aufwands bei der Erstinstallation und freier Skalierbarkeit im laufenden Betrieb wollen die Vertreter der dateibasierten gegenüber der blockbasierten Speicherressourcenverteilung schließlich das bessere Preis-Leistungs-Verhältnis bieten.

Storage-Cluster oder genauer Clustered Storage lautet die Schlüsseltechnik, mit der NAS-Hersteller diesem Markt einen neuen Charme verleihen wollen. Anstelle eines einzelnen NAS-Kopfes wie bei traditionellen NAS-Servern stellen NAS-Cluster mehrere Zugänge zum zentralen Speicher in das Netzwerk, und zwar je einen pro Knoten. Der Durchsatz vom und zum Speicher skaliert dabei theoretisch – laut Herstellern außerdem praktisch – linear über die Anzahl der Speicherknoten, wenn auch nur agglomeriert. *iX* hat einen NAS-Storage-Cluster des mit Hauptsitz in Seattle ansässigen Herstellers Isilon gründlichen Tests unterzogen.

Minimalanforderung für einen NAS-Cluster mit Storage Appliances der Isilon sind drei baugleiche 1-U- oder 2-U-Module mit je vier oder zwölf SATA-Festplatten und damit einer Speicherkapazität von 2 bis 12 TByte pro Box. Das 1-U-Einstiegsmodell IQ 200 kann maximal 24 Knoten und damit 48 TByte Bruttokapazität zu einem globalen Dateisystem bündeln.

Die 2-U-Module mit jeweils 1,92, 3, 6, 9 oder 12 TByte verkraften maximal 95 gleiche Partner, die drei größeren Modelle IQ 6000, 9000 und 12000 können zusätzliche Verstärkung durch Erweiterungs-JBODs (Just a Bunch of Disks) der EX-Familie nutzen. Das obere Ende der Isilon-Produktlinie bildet der IQ/EX 12000-Cluster mit 96 Basisstationen IQ 12000 und daran per SAS gekoppelten Erweiterungen EX 12000, der stolze 2,3 PByte Bruttokapazität zu einem einzelnen Dateisystem bündelt.

Während die 1-U-Einstiegslösung IQ 200 sämtliche, für die vertikale Verteilung der Daten im Cluster erforderliche Kommunikation über Gigabit Ethernet (GE) abwickeln muss, steht den übrigen Modellen des Herstellers dafür Infiniband 4x zur Verfügung. Bei letzteren 2-U-Appliances sind die Frontend-Ethernet- sowie die Intracluster-Infiniband-Ports und die Netzteile außerdem doppelt ausgelegt, sodass kein Single Point of Failure vorliegt.

Ins *iX*-Labor lieferte Isilon drei „Plattform-Knoten“ vom Typ IQ 6000. Jedes System verfügt über zwei 500-Watt-Netzteile. Die Leistungsaufnahme im Betrieb beträgt laut Hersteller typisch 2 und maximal 3,5 Ampere, die Abwärmeleistung beziffert Isilon mit 2400 BTU/h (British Thermal Units per hour). Montage- und Einbauanleitungen waren Bestandteil der Teststellung, das Handbuch stand nur als PDF zur Verfügung. Das auf FreeBSD-Wurzeln

basierende Betriebssystem OneFS war in der Version 4.7.1.8 installiert.

Basis der IQ 6000 Appliances bildet ein Tyan-5360-Mainboard, bestückt mit zwei 3,2 GHz schnellen Xeon-CPUs und 4 GByte RAM. Weitere 512 MByte NVRAM dienen der sicheren Verwahrung des Journals zum Dateisystem, sodass bei Bedarf ein komplettes Zurücksetzen der Hardware auf den ersten Auslieferungszustand ab Werk gelingt. Über den 36-Port-SAS-Edge-Expander VSC7154C von Vitesse steuert ein LSI-Logic-SAS1068-Controller die 12 SATA-Platten und stellt zugleich über drei externe SAS-4x-Ports die Schnittstelle für Erweiterungs-JBODs bereit.

Untereinander kommunizieren die Nodes über die Infiniband-4x-HCAs (Host Channel Adapter) Mellanox MT23108. Die Kommunikation nach außen steuern onboard befindliche GE-Chips vom Typ Intel e1000. Ein Front-Panel-LCD gibt Auskunft über den Cluster-Status. Zusätzlich erlaubt das LCD über vier als Pfeile ausgelegte Tasten eine menügestützte Konfiguration eines Node.

Logisches Zentrum der Isilon-Produkte bildet das aus eigener Entwicklung stammende und patentierte Betriebssystem OneFS. Letzteres beansprucht für sich, die drei klassischen Verfahren geclusterter Speicher „Zweiwege-Ausfallsicherung“, „aggregierter Namensraum“ und „verteiltes Dateisystem“ in einem einzelnen Produkt zu liefern. Intern erzeugt es ein Software-RAID mit über mehrere Knoten verteilten Daten.

Nach außen präsentiert OneFS die Hardware als unitäres Dateisystem, auf das im Netz befindliche Rechner via CIFS, NFS (v3, UDP oder TCP), FTP oder HTTP zugreifen können. Intern steuert OneFS die Datenverwahrung über alle vorhandenen Knoten: redundant, als Stripe-Set und mit extra gesicherten Paritätsdaten.

OneFS bestückt die vorhandene Plattenkapazität mit 8-KByte-Blöcken. Jeweils 16 davon fasst es zu einem Streifen zusammen und verteilt Dateien mit mehr als 128 KByte Speicherbedarf vertikal im Cluster, also auf verschiedene Knoten. Kleinere Dateien liegen vollständig redundant im Dateisystem.

Mit dieser Herangehensweise ist der Speicher nicht nur gegen den Ausfall einer einzelnen Platte stabil; er garantiert die Datenintegrität auch bei Ausfall eines Storage-Knotens. Für den gesamten Cluster, Verzeichnisse oder Dateien per Flex Protect konfigurierbare Sicherheitsstufen erhöhen den Schutz vor Datenverlust bis zur Stufe N+4, für die mindestens neun Knoten erforderlich sind und die eine Datenintegrität selbst bei gleichzeitigem Ausfall von vier Appliances gewährt.

Eine erste Inbetriebnahme gelang bereits nach wenigen Minuten: Die Wizard-gestützte Grundkonfiguration nach dem ersten Booten beschränkt sich auf die Angabe von Passwörtern für Root und den Administrator, dem gewünschten Systemnamen, der Adressbereiche für internes und externes Netz, Gateway, Nameserver und Uhrzeit nebst Zeitzone. Einzige Voraussetzung: Der serielle Port des ersten Systems ist mit einem Terminal zu verbinden, das eine Transferrate (Baudrate) von 115 200 bps unterstützt. Windows- oder Linux-Notebooks mit Hyperterm oder Minicom sowie serieller Schnittstelle leisten das Gewünschte.

iX-TRACT

- Isilon IQ 6000 ist ein skalierbarer NAS-Cluster für NFS-, CIFS-, HTTP- und FTP-Clients.
- Im Cluster vertikal verteilte Daten gewähren die Datenintegrität auch bei gleichzeitigem Ausfall mehrerer Cluster-Nodes.
- Mit jedem zusätzlich integrierten Knoten wächst der Gesamtdurchsatz linear.



Vom unter dem Infiniband-HCA sitzenden SAS-Board führen 12 SATA-Kabel zur Backplane und drei SAS-4x-Ports nach draußen (Abb. 1).

Prinzip Baukastensatz

Weitere Appliances integriert der Administrator wahlweise auf Seite des Knoten, terminalgestützt über die Kommandozeile oder über das frontseitige LED Control-Panel. Alternativ und dabei komfortabel gelingt die Cluster-Erweiterung auch per HTTPS-Zugriff auf die externe Adresse des bereits konfigurierten Node. Technisch bewirkt das Erweitern um einen Knoten ein Vergrößern des Dateisystems. Im Test beanspruchte diese Prozedur weniger als eine Minute, führte selbstverständlich nicht zur Unterbrechung der in Nutzung befindlichen Dateidienste und hatte keinen spürbaren Einfluss auf den Datendurchsatz.

Auch wenn die Administration jedes einzelnen Systems und übergreifend des gesamten Clusters mit immerhin 143 Kommandos möglich ist, alltäglich benötigte Konfigurationen bewältigt der Systemverwalter über das Web-GUI. Anlaufpunkt dazu ist prinzipiell jede externe Adresse des Clusters. Ergänzend



Die SAS- (rechts mittig) und Infiniband-Boards (rechts oben) sind per Reiser-Card horizontal gelegt und übereinander angeordnet (Abb. 2).

lässt sich das gesamte System auch über eine einzelne virtuelle IP-Adresse ansprechen, dazu aber später mehr.

Cluster, Nodes, File System, Modules und Tools lauten die fünf Kategorien, die als Menü-Überschriften Zugang zu den Konfigurationsmöglichkeiten und den Zustandsanzeigen des Systems liefern. Hervorzuheben ist die integrierte kontextsensitive Hilfefunktion, die zu jedem Bildschirminhalt in einem neuen Fenster die zugehörige Seite des HTML-formatierten Handbuchs zum System öffnet. Das IQ WebHelp genannte Nachschlagewerk verfügt über einen Index nebst Glossar und ist dadurch schon beinahe vergleichbar mit typischen, gemäß compressed-HTML-formatierten CHM-Dokumenten.

Im Bereich der Statusanzeigen hat der Administrator Zugriff auf eine Cluster-Darstellung mit links einem Histogramm zum Netzdurchsatz der letzten Stunde und rechts einer Tabelle mit nach Knoten aufgelisteten aktuellen Durchsätzen sowie vorhandener und genutzter Speicherkapazität. Darunter zeigt ein zweiter Block den Status der Datenintegrität und die zuletzt dazu ausgeführten Aktionen, etwa Start und Ende des jüngsten Autobalance (Verteilen der Daten im Cluster). Eine Tabelle mit Rollbalken zeigt Ereignisse an, etwa abgeschlossene Neukonfigurationen oder Ausfall eines Knotens. Diese kann das System auf Wunsch nach Datum, Knotenkennung oder Text der Nachricht sortieren.

Unterschiedliche Meldeverfahren

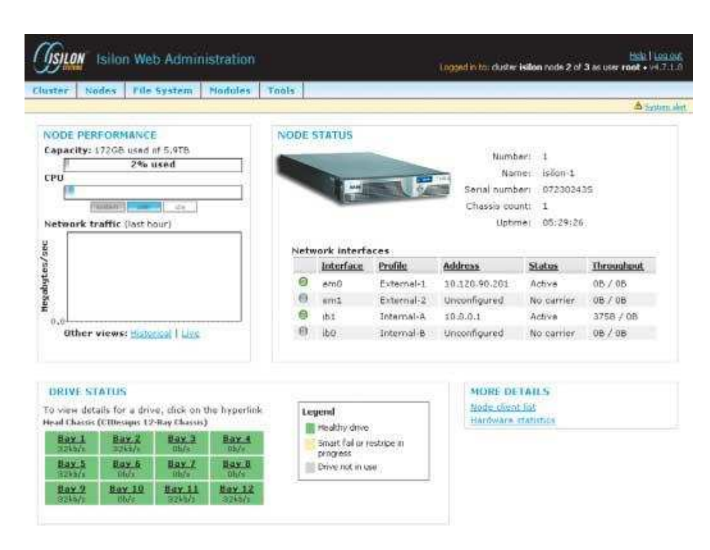
Wer Details zu einer konkreten Appliance einsehen möchte, dem liefern die Unterpunkte des Node-Menüs detaillierte Aussagen. Angezeigte Node-Informationen betreffen den allgemeinen Hardwarestatus der Systemeinheit (Uptime, Hardwaredaten zu CPU, RAM, I/O-Geräten et cetera), Netzdaten (lokale IP-Adressen, Kapazität und Durchsatz) und Status der assemblierten Festplatten. Von dort aus kann der Systemverwalter einzelne Festplatten vom Cluster abmelden, etwa für Wartungszwecke, oder neue Festplatten hinzufügen. Als weiteres Detail erzeugt das System auf Wunsch eine Liste der Clients, die auf diese Knoten CIFS- oder NFS-Shares zugreifen.

Wer mit welcher Zugriffsart Daten vom Cluster beziehen und/oder dort hinterlegen darf, ist im Block der File-System-Menü-Hierarchie anzugeben. Zunächst sind jeweils Grundkonfigurationen für NFS, CIFS, FTP und HTTP einzeln festzulegen. Entsprechend der unterschiedlichen Protokolltypen ist die Anzahl dazu erforderlicher Mindestangaben unterschiedlich: Bei NFS genügt die Festlegung asynchron oder synchron. Für CIFS unterstützte Betriebsarten sind Local-User-, Anonymous- und Domänen-Modus, wobei je nach gewünschter Methode weitere Angaben zu Benutzernamen, Passwort, Dömanenname und so fort zwingend anzugeben sind.

Authentifizierungsschemata kann der Cluster lokal beherbergen oder von einem externen System abrufen, sofern es Namensdienste gemäß NIS, LDAP oder Active Directory bereitstellt. Lokale Benutzer und Gruppen verwaltet das System in BSD-üblicher Manier, und zwar parallel auf jedem Knoten des Clusters mit Einträgen in den zuständigen Dateien *passwd*, *master.passwd* und *group* unterhalb */etc* und automatischem Anlegen eines Heimatverzeichnisses */ifs/home/<user>*. Entsprechend erhalten lokal registrierte Anwender per *ssh* Shell-Zugriff auf jede Appliance – leider auch den Zugang zum Wurzel-Verzeichnis.

NFS-Freigaben erstellt der Administrator unterhalb */ifs*, optional zuzüglich einer Liste *mount*-berechtigter Clients. Voreinstellung ist eine leere Zugriffsliste (jeder darf das Verzeichnis binden) und *rw* als Zugriffsrecht. Eine Mapping-Funktion konvertiert Benutzer-IDs, und zwar entweder für Root oder alle anderen Benutzer. Nobody hat die ID 65534 – RHEL vergibt dafür 99. Im Test trat noch folgender Effekt auf: Per *touch* oder *mkdir* angelegte Dateien oder Ordner erbten stets die Gruppenzugehörigkeit des übergeordneten Verzeichnisses.

Das Erzeugen von CIFS-Freigaben unterstützt ein kleiner *Share Folder Wizard*, der auf bis zu fünf Seiten Namen und Pfad des Shares sowie Zugriffsrechte erfragt. Intern handhabt OneFS den Zugriff durch Windows-Clients über Access-Control-Listen. Der Mechanismus ist Samba-basiert, ergänzt um zahlreiche Zusätze, die für den Einsatz im OneFS-Cluster erforderlich sind.



Zur Verwaltung des gesamten Clusters genügt eine Management-Session zu einem der Nodes. Das auf jeden Knoten anzutreffende Web-Interface gibt Auskunft über den Zustand von Cluster, Nodes, Dateisystem und Modulen und reicht die Konfigurations-befehle weiter (Abb. 3).

Mehrwertdienste optional

Wie bereits erwähnt, stellt die Systemverwalterin die Datenintegrität im Cluster global oder auf Objektebene per Flex Protect bis zum Grad N+4 ein. Die Nettokapazität des Clusters liegt daher stets unterhalb des Bruttovolumens. Auf den drei Knoten der Testlieferung belegten 1 GByte Nutzdaten in der kleinsten Sicherheitsstufe 1,5 GByte Plattenplatz. Auskunft über

den belegten und Umfang der Nutzdaten erzeugt das *du*-Kommando der Gnu-Utilities: *du –k* liefert benutzten Speicherplatz, *du –k ––*apparent-size *die enthaltenen Nutzdaten*.

Mehrwertdienste oberhalb der bisher aufgezeigten Filer-Services von Isilon leisten zusätzliche Softwaremodule, die der Kunde einzeln lizenzieren muss. Das Spektrum der Zusatzfunktionen deckt Anforderungen nach Replikation, Hierarchical Storage Management (HSM), Quotierung, Snapshots und die bereits weiter oben angekündigte Netzwerkvirtualisierung in erweiterter Form ab.

Smart Connect stellt in seiner Grundversion einen DNS-Server, der nach außen eine virtuelle IP-Adresse präsentiert und auf eingehende Anfragen gemäß Round-Robin-Technik die IP-Adresse des jeweils nächsten „freien“ NAS-Kopfes zurückliefert. Smart Connect Advanced berücksichtigt zusätzlich CPU-Auslastung, abgeforderten Durchsatz und Anzahl bereits verbundener Clients bei der Identifikation des „besten“ Kopfes. Dynamisches NFS-Failover und -Failback sowie Load Balancing sind weitere Merkmale des Add-ons.

Geht es um Replikationen, etwa Spiegelung in einen getrennten Brandabschnitt, dann liefert SyncIQ die benötigten Voraussetzungen. Es soll den Kunden beim Disk-to-Disk-Backup, Disaster Recovery und dezentral verteilten Content unterstützen. MigrationIQ hilft beim Aufbau von Multi-Tiered-Speicherhierarchien, indem es automatisiert zu festgelegten Zeiten Dateien oder Verzeichnisse auf einen Hintergrundspeicher migriert.

SnapshotIQ ermöglicht das Anlegen einer laut Hersteller unbegrenzten Anzahl von Snapshots zu Dateien und Verzeichnissen im Dateisystem mit einer Obergrenze von 1024 Snapshots pro Verzeichnis. Neben der traditionellen Copy-on-Write-Technik COW implementiert OneFS zusätzlich eine Point-in-Time-Methode, die bei großen Dateien einen spürbaren Performance-Schub liefern soll. Smart Quotas schließlich ist Isilons Lösung zum benutzerspezifischen Kapazitätsmanagement. Unter anderem benachrichtigt es einen Anwender, dass er seine Kapazität erschöpft hat, stellt ihm aber auch einen Puffer bereit, sodass das Überschreiten der Quota-Grenze kein sofortiges Schreibverbot nach sich zieht.

Laut Isilon leistet ein IQ 6000 basierter Cluster pro NAS-Kopf einen NFS-Durchsatz von 80 MByte/s schreibend und 105 MByte/s lesend bei 3000 I/O-Operationen/s. Im *iX*-Labor durchgeführte *bonnie*-Benchmarks bestätigten diese Werte mit kleinen Abweichungen nach oben und unten. Die Ergebnisse deuten aber darauf hin, dass sie eher die Grenzen der angebundenen NFS-Clients markierten. Speziell etwas ältere PCs mit 32-Bit-CPU und Knoppix, Opensuse oder RHEL 5.1 konnten der IQ 6000 praktisch nur die Hälfte ihrer Leistung entnehmen.

Daten und Preise

3 × IQ 6000 mit je 12 × 500-GByte-SATA-HDs (gesamt 18 TByte raw, 11,64 TByte netto im N+1 Protection Level), 1 × 3,2-GHz-Xeon-CPU, 4 GByte RAM, 512 MByte NVRAM, 2 × GE (Intel e1000), Dual-Port-Infiniband-4x-HCA Mellanox MT23108, 12 × SAS intern und 3 × SAS-4x extern (LSILogic-SAS1068 mit Vitesse VSC7154C)
Preis: 43 502 Euro

3 × OneFS Version 4.7.1.8 mit Autobalance-Software, FlexProtect, NFS, CIFS, HTTP/WebDav, FTP, NDMP
Preis: 38 542 Euro

1 × IQSpare-Kit mit 500-GByte-SATA-HD
Preis: 3350 Euro

2 × Flextronics Infiniband Switch mit 8 Ports
Preis: 5003 Euro

6 × 1-Meter-CX4-Kabel für Infiniband
Preis: 763 Euro

Fazit

Isilon liefert mit der IQ 6000 eine skalierbare Clustered-Storage-Lösung für NAS-Anwender, die aufgrund ihrer Merkmale ein performantes Produkt für große unstrukturierte Datenmengen bildet. Pro Single-Stream handhabt das System NFS-Durchsätze bis zur theoretischen GBit-Übertragungsrate, und zwar an jedem NAS-Kopf. Auch wenn es die Erwartungen I/O-durstiger HPC-Anwender nicht direkt befriedigen kann, im Einsatz für zentralen Dateiservice mit vielen angebundenen Clients zeigt es messbare Vorteile. (sun)

DR. FRED HANTELMANN

ist als IT-Architekt bei der Online Systemhaus ES+C GmbH tätig.

iX-Wertung

⊕ schnelle Konfiguration
⊕ einfache Administration
⊕ freie Skalierbarkeit
⊕ hoher agglomerierter Durchsatz

⊖ schwierige Kapazitätsplanung



4K-Digitalkino-Installation im Cineplex in Münster

Quadratur des Kinos

Dieter Michel





HDTV ist für zu Hause, 4K kommt ins Kino. Die Bildqualität des neuen 4K-Digitalkino-Formats sorgt dafür, dass die Besucher des Cineplex in Münster auf allen Plätzen im Saal ein gleichbleibend hochaufgelöstes Bild ohne Artefakte sehen.

Heimkino ist heutzutage in aller Munde, speziell seitdem hochauflösende Wiedergabegeräte in erschwingliche Preisregionen gekommen sind. „HD Ready“ ist bei LCD- und Plasma-Fernsehern eigentlich schon wieder out – heute muss es bei Neuanschaffungen schon „Full-HD“ sein, also eine Auflösung von 1920 × 1080 Bildpunkten. Hochauflösende Projektoren für den Heimkinobereich sind in bezahlbaren Ausführungen ebenfalls zu haben, und mit der Blu-ray Disc und HD-DVD stehen auch passende Distributionsmedien für Kinofilme zur Verfügung, sodass Cineasten das heimische Wohnzimmer angemessen für den Kinoabend ausstatten können.

Fragt sich, ob überhaupt noch jemand ins Kino geht, und wenn ja, warum eigentlich? Bietet es außer Popcorn einen qualitativen Vorteil gegenüber der bequemen Heimkino-Installation? Diese ketzerischen Fragen kann man durchaus mit „ja“ beantworten, selbst wenn gelegentlich Kinovorstellungen mit unscharf projizierten 35-mm-Kopien in schlechter Qualität einen manchmal zweifeln lassen.

Aber es gibt ja das Digitalkino, bei dem keine Filmkopien klassischen Zuschnitts mehr Verwendung finden. In diesem Bereich hat Sony kürzlich einen spürbaren Schritt nach vorne getan und im Cineplex in Münster eines der ersten Digitalkino-Projektionssysteme in 4K-Auflösung installiert. Zur Einweihung der Installation hatten Sony und die Cineplex-Betreiber Mitte November zu einer Pressepräsentation eingeladen, in deren Rahmen der Hersteller nicht nur das CineAlta-4K-Projektionssystem vorstellte, sondern es den Teilnehmern auch in einem etwa halbstündigen Screening ermöglichte, die Bildqualität des neuen 4K-Digitalkino-Formats im Vergleich mit der bisherigen 2k-Techik sowie analogem Filmmaterial zu sehen und zu beurteilen.

Gründe für einen neuen Standard

Digitale Kino-Projektionssysteme gibt es schon seit einigen Jahren. Derzeit ist in diesem Bereich das sogenannte 2K-Format mit einer Auflösung von 2048 × 1080 Bildpunkten Standard. Eine Weiterentwicklung dieser Norm hat nicht nur ein allgemeines „höher, schneller, weiter“ – also eine Verbesserung um der Verbesserung selbst willen – zum Anlass, sondern wird maßgeblich von zwei Faktoren angetrieben.

Zum einen bietet der bisherige 2K-Digitalkinostandard eine kaum höhere Auflösung als das digitale Fernsehen HDTV. Im Vergleich präsentiert das 2K-Format ein um gerade 128 Pixel breiteres Bild als HDTV bei gleicher Bildhöhe in Pixeln. Für viele Kinofilme mag das zwar in der Praxis ausreichen, ein schlagendes Argument für den Gang ins Kino anstelle des heimischen HTDV-Guckens ist das allerdings nicht – zumal in Zukunft immer mehr potenzielle Kinogänger ein HTDV-Gerät ihr Eigen nennen werden.

Darüber hinaus hat die Digital Cinema Initiative (DCI), eine Vereinigung mehrerer Hollywood-Filmstudios (www.dcimovies.com), für die Entwicklung eines neuen Standards für das Digitalkino oberhalb von 2K handfeste Gründe. Diese beziehen sich sowohl auf die Wahrnehmungsfähigkeiten des menschlichen Auges als auch auf die praktischen Aspekte der hochwertigen Wiedergabe der verschiedenen, kinoüblichen Breitbildformate.

Der wahrnehmungsbezogene Ansatz will eine Projektion ohne sichtbare Pixel ermöglichen. Wer im Kino auf den vom Blickwinkel her optimalen Plätzen – etwa in der Saalmitte oder weiter hinten – sitzt, sieht auch beim gegenwärtigen 2K-Standard keine Digitalisierungsartefakte. Dasselbe gilt übrigens häufig auch für herkömmliche Filmkopien, denn für viele Spezialeffekte dreht man Szenen zwar zum Teil analog auf Film, digitalisiert sie aber im Rahmen der Postproduktion für die Verarbeitung mit CGI-Effekten (Computer Generated Imagery) und belichtet das Ergebnis auf speziellen Kinofilmbelichtern wieder aus – und zwar in 2K-Auflösung. Bei normalem 35-mm-Kinofilm geht damit keine Qualitätsminderung einher, weil hier das 35-mm-Material und der gängige Kinofilmprojektor die qualitätsbegrenzenden Komponenten sind.

Wer bei einer 2K-Projektion im Kino eher weiter vorn sitzt – das konnte man beim Screening in Münster gut

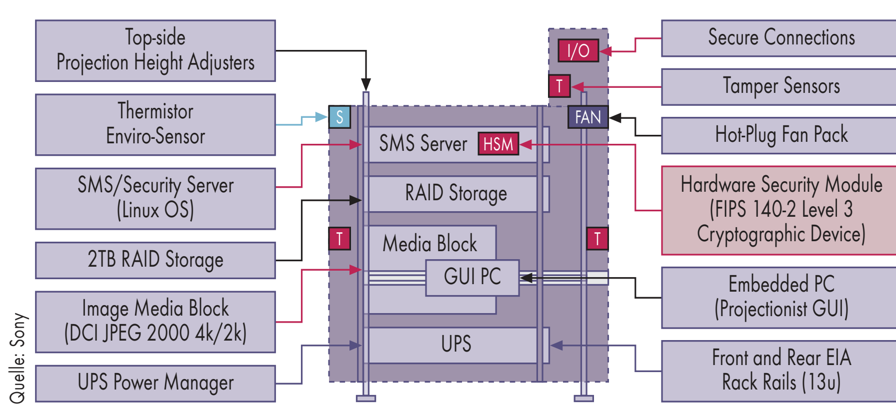
erkennen – kommt zwar immer noch in den Genuss einer sehr guten Bildqualität, speziell im Vergleich mit der analogen Filmkopie. Man kann aber an kontrastreichen Kanten gelegentlich durchaus in geringem Umfang Pixel und Treppchen erkennen, vor allem wenn man im Rahmen eines Qualitätsvergleichs besonders darauf achtet.

Der Ansatz der DCI besteht darin, die Bildauflösung so hoch zu wählen, dass in einem sinnvoll bestuhlten Kinosaal auf praktisch allen Plätzen die Pixel auf der Leinwand so klein erscheinen, dass das menschliche Auge sie nicht mehr als einzelne Bildpunkte wahrnimmt. Das projizierte Bild erscheint dann auf jedem Platz im Kino gleichermaßen scharf.

Als Faustregel für das Auflösungsvermögen des menschlichen Auges gilt, dass zwei benachbarte Objekte nicht mehr separat gesehen werden, wenn der Winkel zwischen ihnen kleiner ist als eine Bogenminute (1/60 Grad). Bezogen auf die Bildhöhe der Leinwand, die eine Kinoprojektion unabhängig vom Seitenverhältnis des Films (Standardformat, Breitwand, Cinemascope et cetera) in der Regel voll ausnutzt, liegt demnach der Betrachtungsabstand, aus dem man die Einzelpixel nicht mehr sieht, beim circa 1,6-Fachen der Bildhöhe. Bei einer für die gebräuchlichen Kinofilmformate bis Cinemascope geeigneten Kinoleinwand mit einer nutzbaren Bildfläche von beispielsweise 15 × 6,3 Meter würde man bei einer 4K-Auflösung bereits aus einem Abstand von 10 Metern die Bildpixel nicht mehr auflösen können.

Projektion in verschiedenen Varianten

In der Praxis ist die nutzbare Sehentfernung noch deutlich geringer, weil man dank der von Sony eingesetzten bildgebenden Projektionstechnik praktisch kein Pixelraster mehr erkennen kann („Fliegengittereffekt“ bei vielen LCD-Projektoren). Sony bezeichnet die beim 4K-Projektor SRX-R110 verwendete Technik als „Silicon X-Tal Reflective Display“ (SXRD). Vom Funktionsprinzip her handelt es sich um ein LCoS-Display (Liquid Crystal on Silicon), das sich im Vergleich zur LCD-Technik durch einen besonders hohen Füllfaktor von 92 % und extrem schmale Übergänge von nur 35 μm zwischen den einzelnen Pixeln bei einem Pixelraster von 8,5 μm auszeichnet. Dies ermöglicht der spezielle Aufbau des reflektiv arbeitenden SXRD-Panels, bei dem die Ansteuerelektronik hinter der Flüssigkristallschicht liegt und nicht wie bei transmissiven LC-Displays im durchleuchteten Bereich des Panels untergebracht werden muss. Daher sind die Pixelübergänge im projizierten Bild kaum sichtbar.



Neben einem genügend leistungsfähigen Projektor benötigt man für die Vorführung eines digitalen Kinofilms diverse weitere Komponenten, die aus Sicherheitsgründen in einer abgeschlossenen Einheit untergebracht sind (Abb. 1).

Die native Auflösung eines 4K-Projektors beträgt 4096 × 2160 Bildpunkte. Das sind nicht doppelt, sondern gleich rund viermal so viele Pixel wie bei 2K. 4K ist demnach nicht 2 × 2K, sondern 2K zum Quadrat – in der Fotobranche würde man da von 8,8 Megapixeln sprechen. Der Vorteil der hohen Auflösung ist, dass auch die Wiedergabe von Filmen im Cinemascope-Format ohne Abstriche bei der Detailtreue erfolgt.

Cinemascope hat ein Seitenverhältnis von 2,39:1 – bei voller Ausnutzung der Bildbreite zeigt der 4K-Projektor 4096 × 1714 Bildpunkte, der 2K-Projektor dagegen nur 2048 × 857. Zum Vergleich: Das normale PAL-Videoformat hat 576 Linien (Bildformat 768 × 576), da sind 857 Linien nicht so furchtbar viel mehr.

Das 4K-Format bietet also nicht eine insgesamt höhere Pixelzahl, sondern verfügt speziell bei Breitbildformaten über mehr Reserven bei der Auflösung. Die DCI-Spezifikation sieht übrigens vor, dass sämtliche Bildformate so skaliert werden, dass sie das Panel-Format in mindestens einer Richtung voll ausnutzen. Damit das Bild immer genau auf die Leinwand passt, muss der Projektor ein motorisch gesteuertes Zoomobjektiv haben, das die Bildhöhe automatisch passend zum Bildformat einstellt – die Bildbreite begrenzen dann die im Kino ohnehin vorhandenen verstellbaren Vorhänge (Caches).

Wie der Film in den Projektor kommt

Die entscheidende Frage beim Digitalkino ist: Wie kommt der Film in den Projektor? Durch die Digitalisierung spart man die teuren und unhandlichen Filmkopien, die in mehreren Teilen angeliefert und für die Vorführung erst zusammengeklebt werden müssen.

Beim digitalen Kino kommt der Film auf Datenträger, und zwar in der Regel auf einer Wechselfestplatte in einem gut geschützten Transportköfferchen. Von dieser Disc kann aber nicht einfach so vorgeführt werden, denn erstens muss das Speichermedium die benötigten Datenraten hergeben, und zweitens ist der Film verschlüsselt.

Für die Filmstudios und mithin das DCI-Konsortium ist besonders der zweite Punkt von Bedeutung, denn

iX-TRACT

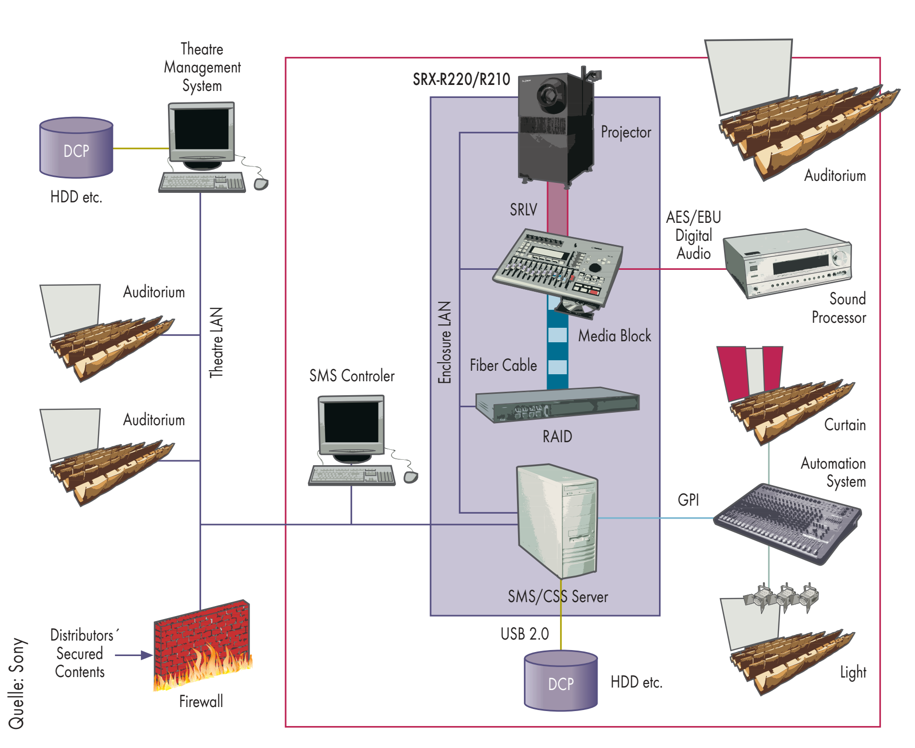
- Die Betreiber des Münsteraner Cineplex-Kinos haben seit November Sonys Digitalkino-Projektionssystem SRX-R110 installiert.
- Der hochauflösende 4K-Projektor sorgt dafür, dass die Bildqualität gleichbleibend gut ist, egal von welchem Sitz aus der Kinobesucher den Film sieht.
- Bei dem System handelt es sich nicht um einen Projektor im herkömmlichen Sinne, sondern eher um einen Computer mit Objektiv. Unter anderem enthält dieser einen Einsteckplatz für eine Festplatte mit dem verschlüsselten Film.

nichts lässt sich leichter und schneller kopieren als eine Festplatte – dafür ist sie ja geradezu gemacht. Ohne weitere Schutzmaßnahmen könnten Unbefugte schnell mal eine Kopie ziehen und hätten quasi ein Master zur Verfügung, auf dessen Basis sich problemlos eine illegale „Zweitverwertung“ aufbauen ließe.

Aus diesem Grund setzt man für die Kodierung der Video- und Audiodaten auf eine starke Verschlüsselung nach FIPS 140-2 Level 3. Die amerikanischen Federal Information Processing Standards (FIPS) definieren Voraussetzungen, die Sicherheitslösungen erfüllen müssen, um für den Schutz vertraulicher Daten in Regierungsbehörden der USA und Kanada zugelassen zu werden. FIPS 140-2 beschreibt Anforderungen an Kryptomodule, die zum Beispiel bei VPN-Lösungen oder Chipkarten zum Einsatz kommen. Definiert werden insgesamt vier Sicherheitsniveaus, von denen der erwähnte Level 3 auch die Manipulationssicherheit des physischen Zugriffs fordert.

Für die Projektion eines digitalen Kinofilms benötigt man also nicht nur einen passenden Projektor, sondern außerdem einen hinreichend großen und schnellen (Festplatten-)Datenspeicher, ein Entschlüsselungsmodul für die Dekodierung der verschlüsselten Festplatteninhalte, das den Anforderungen der FIPS 140-2 Level 3 genügt, sowie einen Image Media Block (IMB) mit JPEG2000-Decoder für die komprimierten Videodaten. Am Ende dieser Verarbeitungskette wird dem Projektor das entschlüsselte Datensignal über eine LVDS-Verbindung (Low Voltage Differential Signaling) zugeführt. Dabei gilt es zu verhindern, das jemand den Datenstrom irgendwie umleitet. Der IMB ist darum auch für die Dekodierung der verschlüsselten Audiodaten sowie eine forensische Markierung zuständig, mit deren Hilfe von der Leinwand abgefilmte Kopien eindeutig einem Serverzertifikat, mithin einem bestimmten Projektor und einer Tageszeit, zuzuordnen sind.

Aufgrund der hohen Sicherheitsanforderungen ist das gesamte Projektionssystem als abgeschlossene Einheit ausgelegt. In dem darin untergebrachten Serverrack befinden sich der Image Media Block, ein Screen Management System (SMS)/Sicherheitsserver auf Linux-Basis mit Hardware-Sicherheitsmodul entsprechend FIPS 140-2, ein 2 TB-RAID-Array, ein Embedded PC für das GUI des Filmvorführers sowie



Eine auf einem externen Rechner installierte Software koordiniert die Kinovorführung und kombiniert den Hauptfilm mit Trailern und Werbung (Abb. 2).

eine UPS (Abbildung 1). Das Gehäuse ist mit Sensoren ausgestattet, die Manipulationen erkennen und bei einem unberechtigten Zugriff, etwa durch Öffnen des Gehäuses, alle gespeicherten Schlüssel löschen.

Für die Ablaufplanung einer Vorführung, beispielsweise die Kombination des Hauptfilms mit Trailern und Werbeclips, ist eine Software auf einem externen Rechner zuständig, der mit dem Projektor vernetzt ist (Abbildung 2).

Computer mit Objektiven

Ergänzend zu dem in Münster installierten Projektionssystem stellte Sony auf der IBC zwei weitere 4K-Kinoprojektoren vor, die nicht den Sony SRX-R110 als Projektor nutzen, sondern eine speziell für den Kinoeinsatz konzipierte Projektionseinheit verwenden. Diese arbeitet mit „nackten“ Xenonkolben, wie sie klassische Kinoprojektoren verwenden, und kann so die Betriebskosten senken. Die beiden neuen Projektoren der SRX-R200-Serie enthalten im Prinzip dieselben IT-Komponenten wie das in Münster installierte System und sehen noch ein bisschen mehr aus wie ein Computer mit Objektiv.

Nachdem die Wiedergabeseite geklärt ist, stellt sich die Frage, woher die neuen, schönen 4K-Digital-Kinofilme kommen sollen. Standard ist bisher 2K, das die Bildqualität einer guten 35-mm-Kopie problemlos wiedergeben kann. Für die 4K-Auflösung reicht die Qualität des 35-mm-Films nicht mehr aus, 4K-Filme werden daher entweder von 65-mm-Kinofilmen abgetastet oder gleich in 4K-Auflösung produziert. Es gibt bereits einige Hersteller (Red, Dalsa und demnächst Sony), die entsprechende Kameras anbieten, sowie Postproduction-Systeme, die mit dieser Auflösung umgehen können.

Fazit

Mit dem zunächst für ein Jahr geplanten Testbetrieb des neuen 4K-Projektionssystems im Cineplex in Münster hat das digitale Kino einen spürbaren und vor allen Dingen sichtbaren Schritt nach vorne getan und bietet neben der großen Leinwand und dem Kinoambiente nun auch von der Bildqualität her einen spür- und sichtbaren Mehrwert gegenüber dem sich zunehmend etablierenden HDTV-Standard. Es lohnt sich, die Schritte wieder einmal ins Kino zu lenken. (ka)

DIETER MICHEL

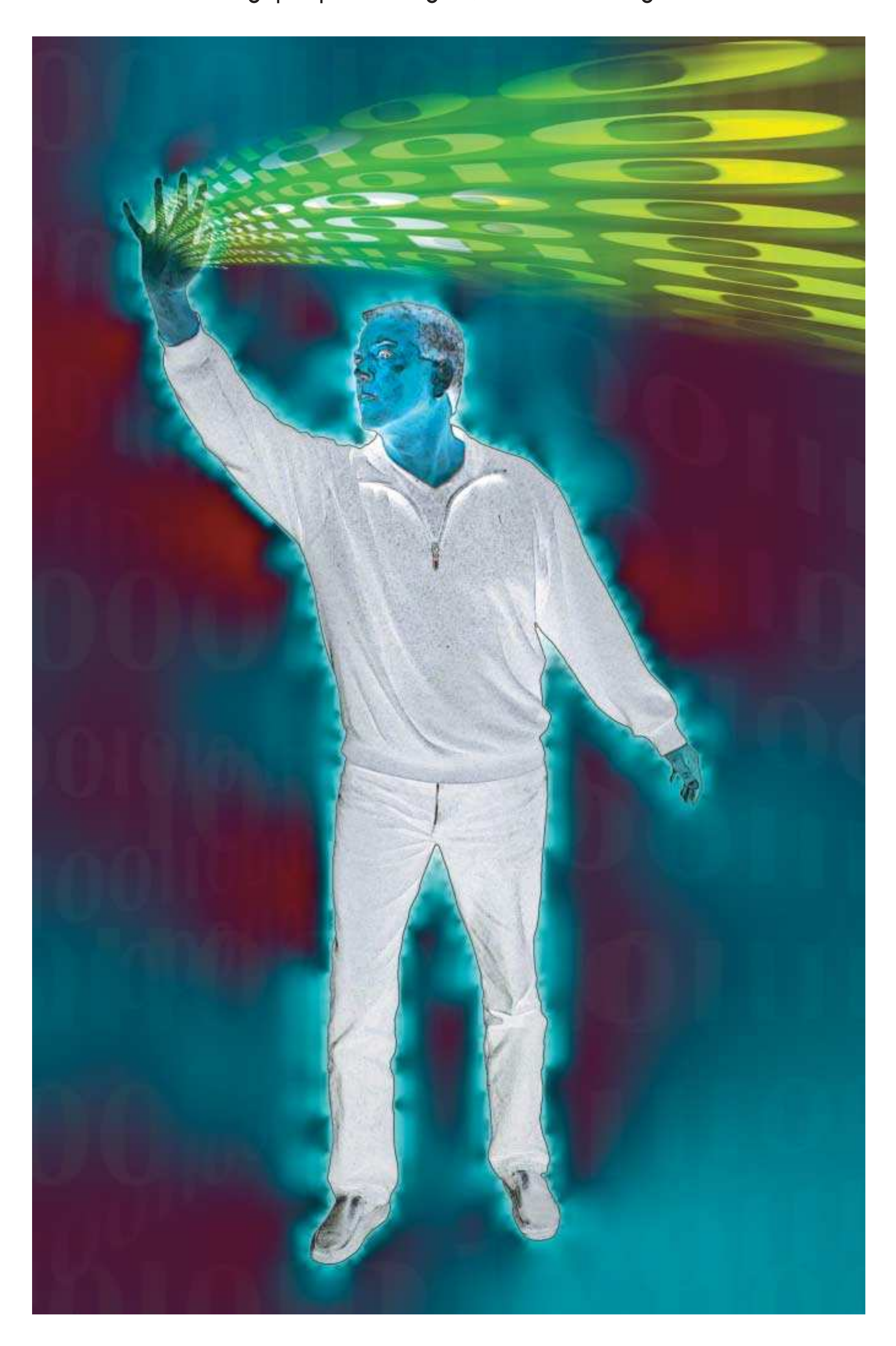
arbeitet als freier DV-Journalist und ist Chefredakteur der Fachzeitschrift Prosound.

Steuerung per Gehirn, Sprache oder Gesten

Bit-schnittig

Barbara Lange

Klassische Eingabemedien sind nicht alles. Überall forschen Wissenschaftler daran, die Schnittstelle zwischen Mensch und Maschine möglichst natürlich zu gestalten. Diverse Forschungsprojekte zeigen, was bereits geht.



Dass sich Computer nicht nur über Tasten oder Knöpfe bedienen lassen, hat spätestens Nintendo mit der Spielekonsole Wii bewiesen. Deren Controller misst über Sensoren die Bewegungen des Anwenders im Raum und überträgt sie auf die digitalen Spielfiguren. Neben Computerspielen existieren diverse Anwendungsbereiche, in denen es sinnvoll wäre, mit dem Rechner intuitiv über die Sinne zu kommunizieren. Zum Beispiel per Sprache oder mit neuen Display-Techniken.

Oder mit der Kraft der Gedanken – diesen Zugangskanal nutzen Brain Computer Interfaces (BCI), die schon seit Jahren erforscht werden [1], [2]. Sie eröffnen damit ganz neue Möglichkeiten in der Kommunikation zwischen Mensch und Maschine. Der Ausgangspunkt: Schon der Gedanke an ein Verhalten verändert die Hirnaktivität, erzeugt messbare mentale Zustände. Wenn jemand denkt „Cursor nach links bewegen“, unterscheidet sich dieser mentale Zustand vom vorherigen, lässt sich selektieren und für die Steuerung von Rechnern verwenden: Für einfache Vorgänge wie Cursor-Bewegungen nach rechts oder links oder für Ja-Nein-Entscheidungen.

Auf dieser Basis arbeiten Brain Computer Interfaces, die die Hirnströme eines Probanden mit einem Elektroenzephalogramm (EEG) aufnehmen, verstärken und in Echtzeit in technische Steuersignale umwandeln. Am sichersten funktioniert dies derzeit bei einer Auswahl von zwei bis drei Alternativen.

Roboterarm per Gedanken steuern

So zeigte das Fraunhofer-Institut für Rechnerarchitektur und Softwaretechnik FIRST auf der Medica 2007 Mitte November in Düsseldorf zum ersten Mal das Projekt Brain2Robot, bei dem ein Proband einen Roboterarm mental steuerte.

Mit diesem Projekt wollten die Wissenschaftler zeigen, dass man einen Roboter ausschließlich auf der Basis von EEG-Messungen führen kann. Ein anderes Verfahren setzen die Neuroprothetiker ein, die im operativen Verfahren Elektroden in das Gehirn implantieren. Mit diesen invasiven Methoden lassen sich zwar 3D-Steuerungssignale erzeugen, gleichzeitig sind aber sowohl die Operation als auch die anschließende

Lernphase mit Versuchstieren, meist Affen, sehr aufwendig.

Konnten Besucher der CeBIT 2006 mithilfe des BCI auf einer „mentalen Schreibmaschine“ Buchstaben mit der Kraft ihrer Gedanken zu Wörtern zusammenfügen [3], so kombiniert Brain2Robot das Brain Computer Interface zusätzlich mit einem Gazetracker, einer Kamera, die die Blickrichtung der Versuchsperson analysiert und erkennt, was sie tun möchte. Schaut sie beispielsweise auf eine Kaffeetasse, schlägt ihr die Software das vor, was man mit einer Kaffeetasse tun kann, etwa: anheben, trinken, Zucker einfüllen. Diese Alternativen bietet ihr das System als Auswahl-Option auf dem Bildschirm an.

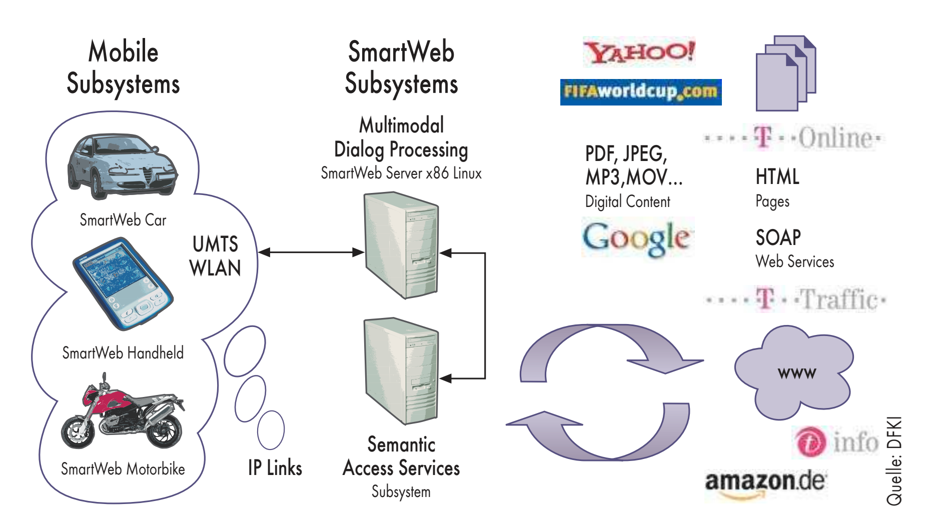
Anschließend setzt das BCI ein: Die 128 Elektroden in der badekappenähnlichen Haube auf dem Kopf der Testperson messen ihre Hirnströme, während sie sich mit der Kraft ihrer Gedanken für eine der angebotenen Aktionen des „intelligenten Schalters“ entscheidet. Das BCI wandelt die Hirnströme in Steuersignale um, der Roboterarm bewegt sich gedankengemäß.

Anwendungsbereiche jenseits der Medizin

Eine Herausforderung ist es dabei, die richtigen Signale aus dem gesamten „Gehirnrauschen“ heraus zu erkennen, betont Prof. Dr. Klaus-Robert Müller, Abteilungsleiter für intelligente Datenanalyse beim Fraunhofer-Institut FIRST. Im Unterschied zu den langen Lernphasen, die die Patienten der Neuroprothetiker auf sich nehmen müssen, um rechnertaugliche Gehirnsignale zu erzeugen, geht dies im gemeinsam mit der Klinik für Neurologie der Berliner Charité entwickelten BCI schnell.

Auch erschließen sich damit neue Anwendungsbereiche jenseits der medizinischen Rehabilitation: Interessant wäre ein mit EEG-Strömen arbeitendes BCI beispielsweise für die Spiele- und Automobilindustrie. Hier sind aber weitere Ideen gefragt, denn die Nutzungsmöglichkeiten dieses neuen zusätzlichen Kanals in der Interaktion von Mensch und Computer sind noch nicht konsequent erschlossen.

Ungelöst ist derzeit noch die Frage, warum ungefähr ein Drittel der Bevölkerung nicht mit dem BCI kommunizieren kann. Ein weiterer Forschungsschwerpunkt betrifft die Sensorik: Müssen die in die Haube eingesetzten Elektroden derzeit mit einem elektrisch gleitenden Gel am Kopf befestigt werden, sollen trocken arbeitende Elektroden eine nächste benutzerfreundlichere Phase einläuten.



Die drei großen Softwarebereiche von SmartWeb für Fußgänger, Auto- und Motorradfahrer. Nach der multimodalen Dialogverarbeitung entscheiden die semantischen Zugriffsdienste, wo das System suchen soll (Abb. 1).

Die Alltagssprache als Weltneuheit

Eine Kombination mehrerer Zugangskanäle nutzt das Projekt SmartWeb, das natürlichsprachliche Fragen wie „Wo geht es zum nächsten Italiener?“ oder „Wer hat bei den Salzburger Festspielen letztes Jahr in der Premiere La Traviata gesungen?“ erlaubt. Besonders mobile Anwendungen sind prädestiniert für intuitive Schnittstellen, die menschliche Äußerungen in Abfragen für Suchmaschinen oder Navigationssysteme verwandeln.

Das Projekt SmartWeb verbindet Sprache, Gestik, Bilder und Video bei der Eingabe, bei der Ausgabe gibt es zusätzlich Grafik- und Tondokumente. Das Bundesministerium für Bildung und Forschung hat SmartWeb mit 13,7 Millionen Euro gefördert, es gab 16 Projektpartner aus Industrie und Forschung. Die Projektleitung liegt beim Deutschen Forschungsinstitut für Künstliche Intelligenz DFKI.

Verglichen mit der Wandlung von Gedanken in technische Steuerungssignale erscheint der Zugangskanal Sprache beinahe trivial. Ist er aber nicht, denn mobile Nutzer im Auto oder Motorrad sowie Fußgänger können Dialoge in Alltagssprache führen, erklärt Prof. Dr. Wolfgang Wahlster vom DFKI, Projektleiter von SmartWeb, und das sei eine echte Weltneuheit.

Auf den mobilen Bildschirmen erscheinen nicht wie üblich lange Ergebnislisten einer Suchmaschine, sondern schlichte Antworten in Text oder Bild, zum Beispiel „Dorfstr. 5“. Bei der anschließenden Frage „Wie komme ich dorthin und muss ich vorher noch tanken?“ berücksichtigt SmartWeb den Inhalt der vorangegangenen Anfrage und reagiert mit einem Navigationsvorschlag in Bild und Ton, verknüpft also Fahrzeugdienste mit Internet-Services.

SmartWeb kombiniert Webdienste

So will man von einer traditionellen Suchmaschine zu einer „Antwortmaschine“ kommen, die das Ergebnis unterschiedlicher Abfragen in einer einzigen Antwort kombiniert, heißt es im DFKI. Die Grundlagen des SmartWeb sollen in das Suchmaschinenprojekt Theseus einfließen, ein vom

iX-TRACT

- Die Interaktion zwischen Mensch und Computer ist seit Langem Thema von Forschungsprojekten.
- Aktuelle Studien erproben die Steuerung von Rechnern und Robotern per Gedankenkraft, gesprochener Sprache und Gestik.
- Multitouch-Bildschirme können mehrere Eingaben gleichzeitig verarbeiten und sind bereits marktreif.



Im Projekt Brain2Robot steuert die Testperson einen Roboterarm per Gedankenkraft. Das Brain Computer Interface zeichnet die Hirnaktivität auf und verwandelt sie in technische Steuersignale (Abb. 2).



Mobile Nutzer im Auto oder Motorrad können im Projekt „SmartWeb“ Dialoge in der Alltagssprache führen (Abb. 3).

Bundesministerium für Wirtschaft und Technologie (BMWi) initiiertes Forschungsprogramm, das das Ziel verfolgt, eine neue internetbasierte Wissensinfrastruktur zu entwickeln.

Sprachsysteme werden immer besser

SmartWeb geht damit weiter als das Vorgängerprojekt SmartKom (bis 2003), das nicht offen im Internet, sondern nur mit einer eigens aufgebauten Wissensbasis funktionierte. Als besonders schwierig bei der Verarbeitung eines unbegrenzten Wortschatzes erwies sich die Erkennung unbekannter Wörter, insbesondere Eigennamen. Herkömmliche Spracherkennungssysteme geben dabei meist das Wort aus, das dem unbekannten am nächsten kommt. SmartWeb erkennt unbekannte Worte als „Out of Vocabulary“ (OOV) und zerlegt sie nahezu fehlerfrei in Silben und Buchstabenfolgen. Sprachfärbungen und national bedingte Akzente erlaubt SmartWeb ebenfalls. Wer als deutscher Nutzer englischsprachige Musiktitel nicht perfekt ausspricht, erhält trotzdem den gewünschten Song.

Insgesamt ermittelte das DFKI bei seiner jährlichen Analyse kommerziell verfügbarer Spracherkennungssysteme auch in diesem Jahr eine Leistungssteigerung. Vorgestellt hat das Institut die Ergebnisse auf den jährlichen Voice-Days im Oktober in Bonn, die nach dem Redaktionsschluss für diese Ausgabe stattfanden.

Verbesserungen gibt es demnach bei der Sprachausgabe: Der Nutzer kann sie abbrechen, wenn er beispielsweise schon bei den ersten Worten merkt, dass die Software ihn missverstanden hat. Toleranter sind die Anwendungen mittlerweile auch bei der Interpretation von Füllwörtern und Versprechern. Darüber hinaus arbeiten alle Systeme, die State-of-the-Art-Techniken verwenden, sprecherunabhängig. Das heißt, der Nutzer muss sie nicht mehr trainieren.

Geforscht wird derzeit im Bereich der Sprecherklassifikation. So ist eine Einteilung nach Alter und Geschlecht in vier Kategorien (Kind/Jugendlicher, Erwachsener, Senior, Geschlecht) bereits möglich, ebenso wie eine emotionale Sprachsynthese.

Nach Angaben des DFKI funktioniert die Identifizierung der Landessprache derzeit noch nicht in allen Fällen verlässlich. Lassen sich türkisch, deutsch und englisch sprechende Nutzer gut voneinander unterscheiden und erkennen, ist die Trennschärfe bei Deutsch, Holländisch und Schweizerdeutsch noch nicht ausreichend.

Ausgehend von der Erkenntnis, dass der Mensch kognitiv überfordert ist, wenn er sich auf einen einzelnen Sinneskanal verlässt, kombiniert das SmartWeb mehrere Kanäle multimodal. Zum Beispiel Bild und Ton: So können Touristen ein Gebäude mit der Handy-Kamera fotografieren und mittels Spracheingabe fragen: „Wie komme ich dahin?“.

Multimodal verschiedene Kanäle verbinden

Bilder, die man als Antwort erhalten hat, lassen sich ebenfalls weiter hinterfragen, etwa das Gruppenbild einer Fußballmannschaft: „Wer ist der Dritte von links?“ Die Auskunft erfolgt auf Grundlage der Beschreibung multimedialer Inhalte durch MPEG7.

Der Motorradfahrer hat Mikrofon und Lautsprecher im Helm und kann über haptische Funktionen, die im Lenker eingebaut sind, Dienste aus dem grafischen Menü selektieren. Eine Kamera misst, wie weit der Fahrer vom Mikrofon entfernt ist oder ob er wirklich die gerade übermittelte Information auf dem Bildschirm wahrgenommen hat. Als Sahnehäubchen erhalten Motorradfahrer zusätzlich von vorausfahrenden Autos – via Internet angebunden – Informationen über den Straßenzustand, Ölspuren oder die Gefahr von Aquaplaning.

Das im SmartWeb entwickelte multimodale Dialogsystem soll in das Projekt „SIM-TD“ (Sichere intelligente Mobilität – Testfeld Deutschland) einfließen, das das BMBF und das BMWi ab Ende 2007 fördern wollen.

SmartWeb ist marktreif. Derzeit ist man dabei, die Projektergebnisse kommerziell umzusetzen. Das Sprachdialogsystem für Motorradfahrer wurde erprobt an der BMW K1200 LT und R 1200 RT. Für die A- und R-Klasse von Mercedes gibt es eine einbaufähige Version.

Fingerfertig: Multitouch-Bildschirme

Eine Weiterentwicklung der herkömmlichen Touchscreens bieten Multitouch-Bildschirme, die mehrere Eingaben gleichzeitig verarbeiten können – zweihändig von einer Person oder von mehreren, die Fenster und grafische Objekte

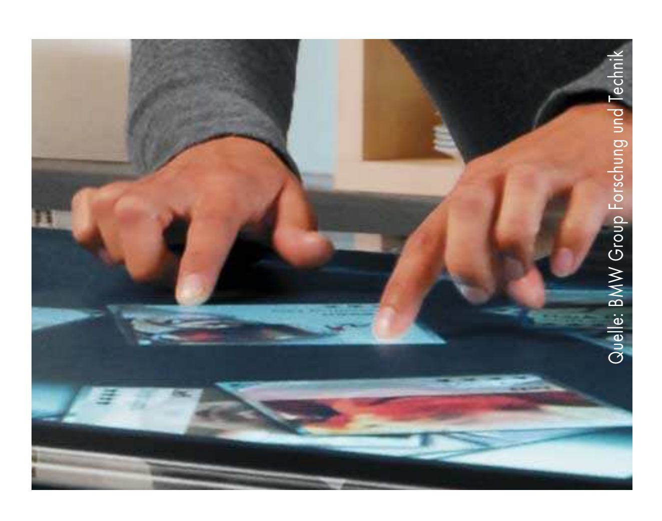
hin- und herschieben, Fenster groß- und kleinziehen, ähnlich wie im Film „Minority Report“.

Bekannt wurde diese intuitiv einsetzbare Technik durch eine Präsentation von Jeff Han von der New York University auf der TED-Konferenz 2006. Der Link für das dort gedrehte Video findet sich im Kasten „Onlinequellen“. Mittlerweile ist die Technik marktreif. So verwendet zum Beispiel die Firma Apple in ihrem neuen, hierzulande im November auf dem Markt erschienen iPhone ein Multitouch-Interface.

Auch Microsoft zielt mit dem für Ende 2007 avisierten Multitouch-Bildschirm „Microsoft Surface“ in Richtung intuitive Benutzerführung. Den 30-Zoll großen Bildschirm steuert der Anwender allein über die Finger. Einsatzgebiete sollen zunächst Restaurants, Hotels und Casinos sein.

Auf der IFA 2007 haben die beiden Unternehmen foresee und werk5 erstmals ihren „Interactive Table“ präsentiert. Als Anwendungsgebiet stellen sich die Entwickler die Präsentation von Projekten vor, da an dem Tisch mehrere Personen gleichzeitig interagieren können. Wo im klassischen Umfeld nur einer an der Maus sitzt, können so mehrere gemeinsam an einer Präsentation arbeiten.

Als besondere Herausforderung bei der Entwicklung des Multitouch-Tisches haben die Unternehmen den Lichteinfall bei Tage gebändigt, daher funktioniert er auch bei Tageslicht. Das Verfahren ist patentiert.



Der Multitouch-Bildschirm „Microsoft Surface“ soll Ende des Jahres auf den Markt kommen. Er kann mehrere Eingabepunkte gleichzeitig verarbeiten (Abb. 4).

Der Touchscreen wird touchless

An der Nutzung von Multitouch für virtuelle 3D-Bildschirme arbeiten derzeit die Forscher am Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut HHI [4]. Was von Weitem aussieht wie ein sinnloses Stechen mit dem Finger in die Luft, entpuppt sich bei näherer Betrachtung als ein Klicken mit dem Finger – zurzeit noch einem einzigen – auf eine virtuelle Buttonleiste, die einige Zentimeter vor dem Bildschirm schwebt.

Autostereoskopische Displays ermöglichen eine dreidimensionale Sicht ohne Stereobrille oder Datenhandschuh. Sie schicken zwei Bilder – eins an jedes Auge, sodass ein räumlicher Eindruck entstehen kann. Für die Interaktion mit Rechner und Bildschirm, beispielsweise das Drücken der virtuellen Auswahltasten, nutzen die Forscher am Fraunhofer-Institut HHI die „Finger-Tracking-Technik“.

So kann der Anwender virtuelle 3D-Objekte bei Mixed-Reality-Anwendungen „berühren“, drehen und bedienen, die 20 Zentimeter vor dem Bildschirm zu schweben scheinen. Zusätzlich erfasst eine Kamera, die in die Tischplatte eingelassen ist, die Bewegungen der Finger. Durch ein Force-Feedback-System erhält der Nutzer das Gefühl einer Tastenempfindung. Zur Erhöhung der Bewegungsfreiheit des Nutzers vor dem Bildschirm prüft eine Kamera ständig den Blickwinkel zwischen ihm und dem Bildschirm.

Wissenschaftler am Heinrich-Hertz-Institut arbeiten daran, dass Multitouch-Technik auch berührungslos funktionieren kann. Dabei sind sie technisch nahe dran, die Bewegung der Fingerspitzen beider Hände videobasiert in Echtzeit zu erfassen, erklärt Dr. Siegmund Pastoor, Leiter Human Factor des HHI. Noch unzureichend erforscht seien Interaktionskonzepte, die eine intuitive Nutzung ermöglichen, das heißt gebrauchstaugliche, plausible Lösungen.

Fazit

Aktuelle Projekte zur Überwindung der klassischen Mensch-Maschine-Schnittstelle nutzen unter anderem als Zugangskanäle Gehirn und Sprache sowie haptische und virtuelle Multitouch-Screen- Eingaben. Viel ist noch Forschung, aber alle Projekte nähern sich der Marktreife oder haben sie schon erreicht.

Die Anwendungsbereiche liegen vor allem in mobilen Szenarien und im medizinischen Bereich zur Unterstützung Behinderter. Nach Einschätzung der Forscher sind die Möglichkeiten der neuen Zugangskanäle aber noch längst nicht erschöpfend genutzt. (ka)

BARBARA LANGE

ist IT-Journalistin und Inhaberin des Redaktionsbüros kurz&einfach in Lengede.

Literatur

[1] Markus Krichel; Denk mal einer an …; Gehirnströme als universelles Eingabemedium; *iX* 11/93, S. 42

[2] Markus Krichel; Brainstorming; Die BioMuse-Entwickler im Gespräch über System-Perspektiven; *iX* 11/93, S. 50

[3] Kersten Auel, Christopher Kunz; Gedankenwelten; Neue Wege der Kommunikation und Interaktion; *iX* 5/06, S. 30

[4] Kersten Auel, Christopher Kunz; Mensch trifft Maschine; Von Robotern und 3D-Displays; *iX* 5/04, S. 32

Onlinequellen

Brain Computer Interface (BCI)	www.bbci.de
SmartWeb	www.smartweb-projekt.de
Multitouch-Technik auf der TED-Konferenz 2006	www.youtube.com/watch?v=PLhMVNdplJc
Interactive Table	www.interactive-table.de



PHP-Anwendungen mit Java-Backends verbinden

Kaffeeklatsch

Markus Eisele

Seit sich Unternehmensprozesse immer mehr ins Web verlagern, wächst das Bedürfnis, etablierte Internet-Techniken wie PHP mit Java-Enterprise-Anwendungen zu kombinieren. Standardprodukte für die Verknüpfung sind zwar noch Mangelware, einige Ansätze jedoch vielversprechend.

Wer sein Java-Backend mit dynamischen Webseiten, oft erstellt mit dem populären PHP, verbinden will, kann sich noch nicht aus einer reichen Auswahl an geeigneter Technik oder gar bei zahlreichen ausgereiften Produkten bedienen. Trotzdem lohnt es sich, einiges von dem Vorhandenen näher zu untersuchen.

Eine Ausführungsumgebung für die Skriptsprache PHP läuft je nach Betriebssystem als separater Dienst oder Prozess. Außer den Schnittstellen zu Webservern wie Apaches HTTPD oder Microsofts IIS bietet PHP keine standardisierten Objekt-Interfaces. Grundsätzlich gilt Letzteres auch für eine Java Virtual Machine (JVM). Will man nun PHP- und Java-Programme dazu bringen, miteinander zu reden, müssen geeignete Zwischenstücke her. Die docken sich an den rudimentären Schnittstellen der jeweiligen Prozesse an, ver- und entpacken übertragene Objekte, die die (für den Entwickler unsichtbare) Kommunikation über Prozessgrenzen hinweg abwickeln, und stellen Programmier-Interfaces (APIs) zur Verfügung.

Bei der Suche nach solchen Adaptern stößt man bald auf die PHP Java Extensions. Diese Erweiterungen nutzen das Object Overloading von PHP, um auf Java-Klassen zuzugreifen. Ruft ein PHP-Programm eine Java-Methode an einem Objekt auf, wird zuerst via Java Native Interface (JNI) eine Java Virtual Machine (JVM) erzeugt. Der Rückgabewert lässt sich in den PHP-Seiten anzeigen. Die ersten Gehversuche mit dieser Lösung gestalten sich mühselig. Der Entwickler muss etliche Dinge konfigurieren, installieren, und die Kombination mit JNI verspricht nicht gerade sonderlich stabile Systeme. Unter Last geht diese Konstruktion schnell in die Knie. Schlimmstenfalls startet jeder Request eine komplette JVM. Für solche Experimente sollte ein Server mit großvolumigem Hauptspeicher zur Verfügung stehen. Instanz-Pooling oder andere Wiederverwertung findet hier nicht statt.

Eingebettete Experimente

Ähnliches lässt sich über die PHP Servlet SAPI berichten. Sie bettet allerdings nicht Java in PHP ein, sondern PHP in Java: Ein Java-Webcontainer (etwa Tomcat) bekommt via Servlet eine PHP-Instanz eingebaut. Obwohl das SAPI-Modul auf den Mechanismen der PHP Java Extensions basiert, ist es stabiler und performanter. Vor allem deswegen, weil sich die Servlet Engine um das Pooling und das Recycling der JVMs kümmert und der PHP-Kern lediglich bei Bedarf geladen wird. Für beide Varianten finden sich Pro- und Kontra-Stimmen in einschlägigen Foren und Artikeln. Gefühlt überwiegen die Problemberichte. Aus diesem Grund sind die PHP Extension für PHP5 als experimentell gekennzeichnet. Für produktive Umgebungen eignen sich beide Ansätze noch nicht.

Dem Java-Entwickler mögen die beiden Vorschläge befremdlich erscheinen, weiß er doch, dass mit Java6 der Java Specification Request 223 umgesetzt wurde. Dieser JSR beschreibt eine API, die einen Weg zum Einbinden von Skriptsprachen in die Java-Welt weist. Pro Sprache definiert der JSR eine sogenannte Scripting Engine. Zwar fördert die Suche in Suns Scripting-Projekt keine entsprechende Engine zutage, woanders findet man jedoch zwei Produkte, die die Scripting API für PHP umsetzen: Die PHP/Java Bridge auf Sourceforge und die native Java-Implementierung von PHP namens Quercus von Caucho. Die API reduziert den Gebrauch von PHP-Skripten in Java zum Vierzeiler (Listing 1).

Damit kann der Java-Entwickler auf einfache Weise Rückgabewerte von beliebigen PHP-Methoden und -Klassen in seine Programme übernehmen. Der Weg in die andere Richtung bleibt jedoch zunächst verschlossen. Allein ist die Scripting Engine nicht in der Lage, PHP-Dateien und -Klassen in einem Java-Applikationsserver auszuführen.

Listing 1

```
ScriptEngineManager m = new ScriptEngineManager();
ScriptEngine phpEngine = m.getEngineByExtension("php");
ScriptContext context = phpEngine.getContext();
Object php2javaResult = phpEngine.eval("<?php echo \"hello world\";
?>",context);
```

Listing 2

```
<servlet>
    <servlet-name>PHPServlet</servlet-name>
      <servlet-class>
         com.caucho.quercus.servlet.QuercusServlet</servlet-class>
    </servlet>
  <servlet-mapping>
    <servlet-name>PHPServlet</servlet-name>
    <url-pattern>*.php</url-pattern>
  </servlet-mapping>
```

Quercus beziehungsweise die PHP/Java Bridge liefern das notwendige Drumherum, dazu später mehr.

Zu den nativen Möglichkeiten gehört die PHP-Java-Interaktion mithilfe von Webservices. Hierbei geht es allerdings nicht um eine enge Verzahnung beider Welten, sondern lediglich um eine lose Zusammenarbeit. Dennoch hat diese Option ihren Reiz. JEE kommt zwar ohne Probleme mit Webservices klar, auf der PHP-Seite muss man jedoch nachrüsten. Infrage kommen dafür etliche Frameworks, beispielsweise NuSOAP, ein Webservices-Toolkit für PHP. Allerdings müssen die Entwickler für diesen Weg auf beiden Seiten Schnittstellen schaffen und Übereinkünfte bezüglich der Zusammenarbeit treffen. Der Gewinn besteht in einer stabilen Lösung, die beide Seiten möglichst wenig beeinflusst. Allerdings muss man hier mit zwei separaten Infrastrukturen klarkommen.

Offen in alle Richtungen

Die im Folgenden vorgestellten Integrationsbemühungen gehen tiefer. JSR-223 bringt zwar dem Java-Entwickler PHP näher, seine PHP-Kollegen bleiben aber ohne nennenswerte Unterstützung. Dieses Manko wollen verschiedene Brücken beheben. Zur erwähnten PHP/Java Bridge existiert eine kommerzielle Alternative von Zend, der „PHP Company“. Beide Produkte sind in der Lage, von PHP aus auf Plain Old Java Objects (POJOs) sowie andere Java- und JEE-Komponenten (EJBs, Connections, JMS et cetera) zuzugreifen. Die PHP/Java Bridge bietet zwei Zugriffsoptionen: Eine native, in PHP geschriebene Java-Brücke und ein in C umgesetztes PHP-Erweiterungsmodul. Der erste Weg lässt sich einfach beschreiten, man muss nur die relevanten PHP-Klassen auf den Webserver kopieren. Folgender Code-Schnipsel eröffnet den Zugriff auf Java-Klassen:

```
?php  require_once(„java/Java.php“);
  $string = new Java(„java.lang.String“, ⁊
                              „HelloWorld“);
  echo $string;
  ?>
```

Zum Ausführen benötigt diese Variante keinen nativen Code (*.so* oder *.dll*). Großes Manko: Im Gegensatz zum in C verfassten Modul läuft sie rund zehnmal langsamer. Dafür ist die Konfiguration des C-Moduls schwierig. Außer der *JavaBridge.jar* muss der Entwickler eine spezielle *php_java.dll* in den PHP-Erweiterungsordner (*<drive>/php/ext*) kopieren und in (*php.ini*) registrieren. Unter Linux funktioniert das alles ohne installiertes Java; Windows verlangt hingegen ein Java Development Kit (JDK). Das einsatzbereite Modul kann beispielsweise folgenden Code ausführen:

```
<?php
$system = new Java('java.lang.System');
echo 'Java version=' . $system->⁊
                 getProperty('java.version');
?>
```

Interessant ist die Tatsache, dass die PHP/Java Bridge ohne JNI auskommt. Neben den zwei beschriebenen Modulen und der Scripting Engine gibt es noch eine Mono-.Net-Brücke. Detaillierte Informationen dazu stehen auf der Projektwebseite.

Zend bietet im Rahmen seiner PHP-Plattform die Java Integration Bridge an. Leider gibt die Firmenwebsite dazu nicht allzu viel Wissenswertes preis. Erst der User Guide erschließt die Architektur des Produkts. Das Java Middleware Module folgt dem PHP-Standard. Der Zugriff sieht genauso aus wie bei der PHP/Java Bridge. Allerdings helfen hier einige Zusatzprogramme. Sowohl auf PHP- als auch auf Java-Seite wird eine Komponente installiert, die das Ein- und Auspacken der übertragenen Objekte besorgen.

Gute Lösung, etwas gebremst

Um die Wiederverwendung von Java-Instanzen und ein besseres Handling (Pooling, Threading et cetera) zu ge-

währleisten, startet man nicht einfach eine beliebige JVM und konfiguriert PHP zum Zugriff darauf, sondern aktiviert einen speziellen Dienst, der die Java-Instanzen verwaltet. Über diesen Service lassen sich sowohl Port als auch maximal verfügbare Prozesse einstellen. Die Kommunikation mit dem JEE-Applikationsserver erfolgt via Remote Method Invocation (RMI). Nachteil dabei: Ein PHP-Programm kann nur auf Objekte zugreifen, die über ein Remote-Interface verfügen. In Sachen Performance profitiert dieser Weg also nicht von den in der Java-Spezifikation verankerten Optimierungen, den EJB Local Interfaces. Nach eigenen Angaben arbeitet Zend zurzeit an dieser Baustelle. Es sei noch erwähnt, dass die Firma zwar zusammen mit Sun den JSR-223 ins Leben rief, bisher aber keine offizielle PHP-Engine für Java vorgestellt hat.

Wer mit den bisher beschriebenen Möglichkeiten nicht zurechtkommt, dem bleibt eine letzte Option. Neben der offiziellen PHP-Distribution gibt es die nach der GPL lizenzierte und in Java geschriebene Ausführung Quercus. Dieses Wunderwerk stammt, wie oben schon erwähnt, von der Firma Caucho, Hersteller des Applikationsservers Resin, der auch Quercus enthält. Quercus' Scripting Engine ist dieselbe wie bei der PHP/Java Bridge. Vorteile zieht der Entwickler aus den zusätzlichen Funktionen, die das Paket bereitstellt. Dazu zählen ein Servlet Wrapper für PHP-Aufrufe sowie die komplette Verwaltung der PHP-Instanz innerhalb des Webcontainers. Quercus lässt sich als Webapplikation (*.war*) herunterladen. Nach dem Auspacken muss der Programmierer diese Datei in das Anwendungsverzeichnis eines beliebigen Webcontainers schieben und schon steht ihm ein halbwegs aktuelles PHP 5.2.0 unter Java zur Verfügung. Damit ist es denkbar einfach, vorhandenen Java-Anwendungen PHP-Funktionen anzubieten. Zwei Java-Bibliotheken muss er noch in das *WEB-INF/lib* kopieren und das PHP Servlet in der *web.xml* registrieren (Listing 2).

Leider ist auch hier nicht alles so einfach, wie es zunächst scheint. Zwar stehen schon eine Menge PHP-Module für Quercus zur Verfügung (darunter APC, iconv, GD, gettext, JSON, MySQL, Oracle, PDF und Postgres) aber die komplette Palette deckt das Angebot noch nicht ab. Dennoch sind viele bekannte PHP-Anwendungen unter Quercus lauffähig, etwa DokuWiki, Drupal, Gallery2, Joomla, Mambo, Mantis, MediaWiki, Phorum, phpBB, phpMyAdmin, PHP-Nuke, Wordpress und XOOPS. Leider enttäuscht die kostenlose Version des Werkzeugs mit einem großem Nachteil: Datenbankzugriffe lassen sich nur per Java Naming and Directory Interface (JNDI) ausführen. Der normale PHP-Zugriff funktioniert nur in der lizenzpflichtigen Professional-Version des Application Server (Listing 3).

Ebenfalls ein Schmankerl, in dessen Genuss nur die Professional-Lizenznehmer kommen: Quercus kann PHP-Quellen nicht nur zur Laufzeit interpretieren, sondern auf Wunsch auch vorkompilieren. In produktiven Umgebungen ergibt sich hier sicher ein Performancegewinn.

Es ist nicht möglich, eine pauschale Empfehlung für den richtigen Weg zu geben, zu unterschiedlich sind die Anforderungen. Wer in PHP-Webseiten gelegentlich einige Java-Funktionen verwenden möchte, muss eine andere Architektur aufbauen, als jemand, der eine Java-Webanwendung um PHP-Teile erweitern will. Das am häufigsten gewünschte Szenario ist vermutlich das anfangs skizzierte Java-Backend mit Geschäftsdaten im Zusammenspiel mit einer agilen und modernen PHP-Weboberfläche. Bei Unternehmen, die so etwas wollen, kann ein Produkt mit professionellem Support punkten – und hier bleibt nur Zends Java Integration Bridge. Wer nicht auf kommerzielle Unterstützung angewiesen ist, kann sich am PHP/Java-Bridge-Projekt auf Sourceforge versuchen. Am einfachsten gestaltet sich die Kommunikation über Webservices (XMLRPC/SOAP). Dafür braucht man in der Regel nicht einmal eine zusätzliche Architekturkomponente, bekommt allerdings auch keine integrierte Lösung. (jd)

MARKUS EISELE

arbeitet im Bereich Software-Technologie im Center of Competence IT-Architecture der MSG Systems AG.

Listing 3

```
// PHP-DB-Zugriff
//mysql_connect($host, $username, $password, $dbname);
// JNDI-Zugriff
mysql_connect("java:comp/env/jdbc/myDatabaseName");
```

JEE und die Technik

Die JEE-Spezifikation stellt einen allgemein akzeptierten Rahmen bereit, in dem sich aus modularen Komponenten verteilte, mehrschichtige Anwendungen entwickeln lassen. Um eine JEE-Anwendung ausführen und betreiben zu können, benötigt man einen Applikationsserver, der Ausführungsumgebungen für die Komponenten zur Verfügung stellt. Zumeist enthält der Application Server einen Webserver, der das generierte HTML an den Browser ausliefert. Lediglich ein Teil der JEE, nämlich die Java Server Pages (JSP) sind mit PHP-Seiten vergleichbar: Bei beiden handelt es sich um mit Code angereicherte HTML-Seiten. Im Gegensatz zu den PHP-Seiten werden JSPs beim erstmaligen Ausführen in Servlets konvertiert. Ein Servlet ist eine programmierte oder generierte Java-Klasse, die sich in einem JEE-Webcontainer ausführen lässt.

PHP: Technische Grundlagen

Um eine PHP-Datei im Rahmen einer Webanwendung ausführen zu können, benötigt man ein System, das mit den im Sourcecode enthaltenen Anweisungen umgehen kann. Herkömmliche Webserver sind dazu nicht in der Lage. Sie bieten aber eine Schnittstelle (beispielsweise ISAPI oder CGI) für die Interpreter an. Sobald ein Webserver eine registrierte Datei (etwa **.php*) ausliefern soll, übergibt sie ein Server-Daemon oder Dienst (zum Beispiel Apache oder IIS) an den Interpreter. Performant ist das Ganze jedoch nicht. Daher hat man dieses Vorgehen über die Jahre Stück für Stück durch Apache-Module abgelöst. Apaches Webserver bietet heute diverse Komponenten für die Zusammenarbeit mit Skriptsprachen. *mod_php* ist für das Einbinden des PHP-Interpreters zuständig.

Onlinequellen

Zend PHP Java Bridge	www.zend.com/products/zend_platform/in_depth/php_java_integration
PHP-Java-Bridge	php-java-bridge.sourceforge.net/pjb/
JSR 223	www.jcp.org/en/jsr/detail?id=223
Scripting Engines	scripting.dev.java.net/
Eclipse PDT	www.eclipse.org/pdt/
PHP in Java	quercus.caucho.com//
NuSOAP	sourceforge.net/projects/nusoap/

Business Rules Management: Ende des Release-Zyklus

Normalfall Beta

Wolfgang Martin



Geschäftsprozess- und Regelmanagement in einem serviceorientierten IT-Umfeld eröffnen der Softwareentwicklung neue Perspektiven. Die mit viel Aufwand verbundenen Release-Wechsel großer Systeme könnten beispielsweise bald der Vergangenheit angehören.

Prozessorientierung heißt: konsequentes Denken in Geschäftsabläufen und -regeln. Die fachlichen Gegebenheiten bildet ein Modellierer mithilfe eines Business-Process-Management-Werkzeugs (BPM) ab, dessen Kern aus einer sogenannten Prozessmaschine besteht. Mit Regeln beschreibt er die Entscheidungsflüsse, hinter denen sich die prozessübergreifenden Managementprinzipien und die Firmenpolitik verbergen. Für das Formulieren und Ausführen der Regeln gibt es „Regelmaschinen"; die zugehörige Disziplin nennt sich Business Rules Management (BRM).

Ein Prozess nutzt typischerweise mehrere Regeln, während eine Regel in verschiedenen Prozessen ihren Dienst verrichten kann. Es ist daher notwendig, Prozess- und Entscheidungslogik strikt voneinander zu trennen. So erhält das Unternehmen klare Abläufe und die Angestellten verstehen die hauseigenen Prinzipien besser. Nur unter diesen Voraussetzungen kann ein Unternehmen flexibel reagieren. Seine Agilität bezeichnet die Fähigkeit zur Innovation und schnellen Anpassung der Geschäftsmodelle. Von industrialisierten beziehungsweise standardisierten und automatisierten Abläufen erhofft man sich eine kontinuierliche Optimierung der Prozesse sowie geringere Kosten. Ziel ist es, Durchlaufzeiten, -volumen und die Qualität der Produkte ständig zu verbessern.

Eigentlich widersprechen sich Agilität und Industrialisierung. Wenn ein Unternehmen jedoch Prozesse und Regeln im Kontext einer serviceorientierten Architektur (SOA) verwaltet, hat es die Möglichkeit, beide Zielsetzungen mittels BPM und BRM zusammenzubringen. Denn SOA zielt darauf ab, „Software for Change" zu verwirklichen. BPM schafft die technischen Grundlagen für das Automatisieren von Geschäftsprozessen, BRM liefert die Basis für das Standardisieren sowie eine transparente Firmenpolitik und nachvollziehbare Managementprinzipien.

Regeln integrieren sich in die SOA

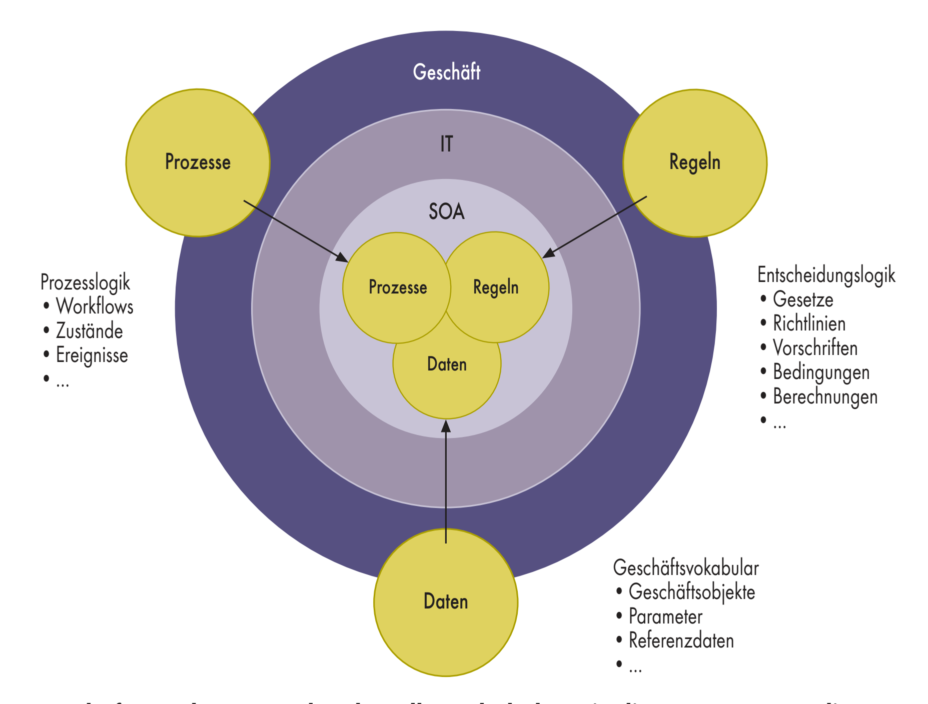
Eine SOA definiert das Zusammenspiel von Prozessen und Regeln neu. Bisher galt das von der Gartner Group aufgestellte Dogma, beide strikt voneinander zu trennen und zu verwalten, die Prozessmaschinen beziehungsweise Regelmaschinen aber miteinander zu integrieren. Die SOA-Idee geht einen Schritt weiter: Statt proprietäre BPM- und BRM-Techniken zusammenzuführen, gelten die Regeln hier als von der Prozessmaschine orchestrierte Dienste. Sie bilden eine neue Kategorie namens Rule-Services, die sich als Kapselung komplexer Regeln verstehen lassen. Ein Rule-Service kann einen anderen aufrufen – analog zum Unterprozess-Prinzip im BPM. Damit wird die Regelmaschine zum Bestandteil der SOA, und die Entscheidungslogik Teilmenge der Geschäftslogik. Für die Regeln existiert eine eigene Administration im Repository, für die Registrierung und in der Governance-Infrastruktur. Die Regeln basieren auf dem gleichen Geschäftsvokabular wie die Prozesse (Abbildung 1).

BPM und BRM sind Kreislaufmodelle, die Prozesse beschreiben, Regeln planen, modellieren, implementieren, überwachen und steuern (Abbildung 2). Hier kommt es nicht nur auf die richtige Architektur an, sondern auch auf die Zusammenarbeit zwischen Fachabteilung und IT. In der Vergangenheit konzentrierte sich diese Kooperation beim Ändern von Softwaresystemen auf den Release-Wechsel. Die Fachabteilung stellte Anforderungen, die IT setzte sie in einer neuen Release um. Das ist selbst im BPM noch gang und gäbe,

obwohl man heute typischerweise streng zwischen dem fachlichen und technischen Design von Prozessmodellen trennt. BRM ist schon weiter: Systeme wie Visual Rules von Innovations bieten Werkzeuge an, mit deren Hilfe Angestellte aus den Fachabteilungen aktiv im BRM-Prozess mitwirken können.

Änderungen an den Regeln sind „die Regel“, wie eine Studie der IDC unterstreicht (Kasten „Nicht in Stein gemeißelt“). Die Möglichkeit, Geschäftsregeln außerhalb der Release-Zyklen schnell an neue Gegebenheiten anzupassen, setzt einen durchgehenden Total-Quality-Management-Prozess (TQM) mit klaren Verantwortlichkeiten voraus. Dieser muss sowohl die Bedürfnisse hinsichtlich der fachlichen Umsetzung der Regeln als auch die technischen Anforderungen der IT hinsichtlich automatisierter Test-, Freigabe- und Deployment-Verfahren erfüllen.

Wichtig ist in diesem Zusammenhang, eine Historie der Regeln zu erstellen und sie im Sinne eines Performance-Management zu messen. Beispiel: Ein Gesetz kommt im März mit Rückwirkung zum Januar und wird als Regel im April eingerichtet. Das bedeutet, dass das Unternehmen es auf den Zeitraum von Januar bis März anwenden kann, aber nicht muss. Nun lässt sich simulieren, für welche Fälle das alte oder das neue Gesetz einen höheren Nutzen bringt.



Geschäftsregeln verwenden dasselbe Vokabular wie die Prozesse, unterliegen aber einem eigenständigen Management (Abb. 1).

Eigenes Leben für Geschäftsregeln

Erst die Unterstützung im gesamten BR-Prozess durch ein BRM-System erlaubt schnelle und sichere Regeländerungen während des laufenden Betriebs. Das gewährleisten insbesondere die weitreichenden Testfunktionen, die automatisch erstellte, stets aktuelle Dokumentation und die einfache Nachvollziehbarkeit der Regeln durch Fach- sowie IT-Abteilung. Als große Schwachstelle im BPM gilt das Repository, an dieser Stelle können die BRM-Anbieter ihre Stärke ausspielen. Fazit: Regeln haben ihren eigenen Lebenszyklus, der ein neues Management jenseits traditioneller Release-Verfahren erfordert.

Praktische Erfahrungen mit diesem Ansatz sprechen für sich (Abbildung 3). Die Post Finance in Bern setzt bei der Umsetzung des Geldwäschegesetzes (GwG) auf Visual Rules von Innovations. Das BRM basiert auf einem definierten Prozess, der die Schritte „Änderungsantrag“, „Bewilligung und Freigabe der Änderung“, „Dokumentation der Regel“ und die Aktivierung umfasst. Die Gesamtverantwortung für den Prozess liegt bei der Fachabteilung Compliance. Hier gibt es eine Gruppe, die sich um die Weiterentwicklung der Compliance-Methoden sowie um die Verbesserung und Kontrolle der benötigten Werkzeuge kümmert. Der Verantwortliche für die Tools ist auch zuständig für den BRM-Prozess.

Die GwG-Szenarien, die per BRM umgesetzt werden sollen, arbeitet er in Zusammenarbeit mit dem Leiter Compliance aus, der auch die formelle Freigabe erteilt. Letzteres soll künftig ein Konzernausschuss erledigen. Das Modellieren und Testen liegt in der alleinigen Obhut des BRM-Verantwortlichen, ebenso wie die Überwachung der aktiven Szenarien. Die IT ist nur gefragt, wenn es um das Einführen neuer Parameter für das Regelwerk geht, und die Testabteilung überprüft bei Release-Wechseln einige Funktionen.

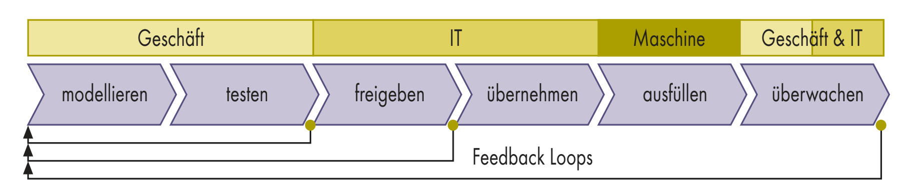
Zusätzlich läuft in der Abteilung Compliance ein Reporting, das Unregelmäßigkeiten im Laufe einer Woche sicher erkennt. Der BRM-Verantwortliche und sein Team betreuen außerdem die operational tätigen Mitarbeiter im Umgang mit den Auswirkungen der Szenarien. Hier lassen sich Unklarheiten und Unschärfen identifizieren. Seit dem Einsatz des Systems im Jahr 2005 sind bisher keine fehlerhaften Konstellationen in den Prozessen entstanden.

Zu Änderungen an den Szenarien kommt es alle zwei bis drei Monate. Dabei vergehen zwischen der Analyse und der Implementierung zwischen einer und drei Wochen. Schneller geht es bei

iX-TRACT

- Prozessorientierte Unternehmen zeichnen sich aus durch konsequentes Denken und Handeln in Geschäftsabläufen und -regeln.
- Durchgänge Geschäftsprozesse sollen einerseits einen hohen Industrialisierungsgrad bewirken, andererseits der Organisation flexiblen Handlungsspielraum verschaffen.
- Business Rules Management und Business Prozess Management in einer serviceorientierten Architektur beeinflussen nicht nur die Geschäftsprozesse, sondern auch die Softwareentwicklung, starre Release-Zyklen können entfallen.

technisch bedingten Anpassungen, die den Wirkmechanismus der Szenarien nicht grundsätzlich ändern, der minimale Implementierungszeitraum kann bis auf einen Tag schrumpfen.



Lebenszyklusmanagement ist ein Prozess, der Regeländerungen im laufenden Betrieb zulässt (Abb. 2).

Unternehmen	Szenario	Einsparung
PostFinance, Schweiz	Transaktionsanalyse, Geldwäsche-Erkennung	25 %
Sungard Futures Systems, USA	Handelssystem für börsennotierte Derivate	50 %
MobileSelect, Australien	Webbasiertes Beratungsportal für Geräte und Tarife	66 %
Sungard Forbatec	Steuerung eines Fond-Management-Systems	70 %

Initialeinsparung	25 % – 50 %
Änderungseinsparung	50 % – 70 %

Mit einem BRM-System können Unternehmen Zeit gegenüber konventioneller Programmierung gewinnen (Abb. 3).

Klassische Verfahren zu starr

Die HypoVereinsbank setzt Visual Rules in einem Frühwarnsystem für Kreditrisiken ein. Täglich analysiert es rund 800 Kapitalmarktadressen. Klassische Verfahren der Softwareentwicklung sind nach Einschätzung der HVB zu starr und zu langsam, um ein Frühwarnsystem an die Marktdynamik anzupassen. Anstatt die Aufgaben aufzuschreiben, sollen die fachlich Verantwortlichen ihre Regeln selbst in Entscheidungsbäumen grafisch modellieren und mit echten Marktdaten testen. Dieser Analyseprozess erleichtert das Feintuning erheblich. Für jeden Regelbaum wird im letzten Schritt der gesamte Programmcode automatisch generiert, dokumentiert, versioniert und in das Produktionssystem transportiert. Das Vorgehensmodell der HVB für dieses Projekt basiert auf zwei Säulen.

Zunächst gibt es eine neue Arbeitsteilung zwischen IT und Fachbereich. Die Fachverantwortlichen entwickeln und warten das Frühwarnsystem selbst; die IT bereitet die Input-Daten als Basis für das Regelsystem auf und kümmert sich um die Integration des generierten Codes in das Produktionssystem.

Weiterhin gilt die sogenannte Continuous Integration als Vorgehensmodell. Regeln lassen sich kurzfristig ergänzen und modifizieren, tägliche Releases ermöglichen die kontinuierliche Weiterentwicklung. Codegenerierung, Dokumentation und Deployment erfolgen automatisch.

Der klassische Software-Release-Zyklus von Konzeption, Realisierung, Test mit begleitender Qualitätssicherung und Dokumentation ändert sich mit dem Einsatz eines BRM-Systems. Da die Beteiligten direkt mit dem Modellieren der Regeln beginnen, verkürzt sich die Konzeptionsphase deutlich. Bereits hier legen sie Testdaten und Referenzergebnisse fest. Dieser Ansatz ermöglicht ein iteratives Vorgehen bei der Definition der Geschäftsregeln (insbesondere bei komplexen Regeln unverzichtbar) und legt die Grundlage für die automatisierten Testverfahren im TQM. Eine Kodierung entfällt komplett, und damit erledigen sich auch die Abstimmungsschleifen zwischen Fachbereich und IT.

Statt eines traditionellen Release-Verfahrens entsteht so ein dynamisches, kontinuierliches Lebenszyklus-Management der Geschäftsregeln. Interessanterweise entspricht dieses Verfahren einem der Grundprinzipien des sogenannten Web 2.0. Eine der Thesen von Tim O'Reilly, der den Begriff mit geprägt hat, besagt, dass der traditionelle Software-Release-Zyklus am Ende ist. Der Open-Source-Anspruch „release early and release often" hat sich seiner Ansicht nach zu einer radikaleren Idee gewandelt: die permanente Beta-Version, in die der Entwickler neue Funktionen auf einer monatlichen, wöchentlichen oder gar täglichen Basis einfügt. BRM erfüllt dieses Paradigma nicht nur – es erlaubt sogar ein kontinuierliches Lebenszyklus-Management in kontrollierter Form als Kollaborationsmodell für Fachbereiche und IT. (jd)

Nicht in Stein gemeißelt

Geschäftsregeln sind naturgemäß nicht für die Ewigkeit gemacht, sie können sich sogar schnell ändern. Hier ein Beispiel aus der Portfolioanalyse einer Privatbank:

„Wenn die Fälligkeiten des Kunden größer als 50 000 Euro sind und die Bank ein Kundenmandat besitzt, dann reinvestiere die Fälligkeiten umgehend."

„Wenn die Fälligkeiten des Kunden größer als 50 000 Euro sind und die Bank kein Kundenmandat besitzt und der Kunde kontaktiert werden kann, dann rufe ihn an."

„Wenn die Fälligkeiten des Kunden größer als 50 000 Euro sind und die Bank kein Kundenmandat besitzt und der Kunde nicht kontaktiert werden kann, dann befrage ihn beim nächsten Kontakt."

„Wenn die Fälligkeiten des Kunden größer als 100 000 Euro sind, dann zeige einen Hinweis auf dem Portal des Kundenberaters an."



DR. WOLFGANG MARTIN

ist ein europäischer Experte auf den Gebieten BI/CPM, BPM, SOA und CRM.



50 Jahre ARPA

Beim Schopf gepackt

Detlef Borchers

Dass das Internet als Reaktion auf den Sputnik-Schock entstand und atomkriegsicher sein sollte, sind zwei immer wieder gern erzählte Geschichten. Die Wirklichkeit ist wie so oft ein wenig komplexer.

Am 4. Oktober 1957 schoss die Sowjetunion um 22:28 Uhr GMT den ersten künstlichen Satelliten in eine Erdumlaufbahn. Sputnik (Gefährte) genannt, besaß dieser Satellit nichts weiter als einen robusten Radiosender, dessen Piepsen Funkamateure in aller Welt empfangen konnten. „Erster!“, signalisierte das Piepsen den Amerikanern.

Drei Wochen lang funkte der Radiosender im Sputnik sein 4 Sekunden langes Piepsen auf den Frequenzen von 20 und 40 MHz. Dann waren die ihn in Form einer Mutter umgebenden Silber-Zink-Batterien leer und die kleine, 83,6 Kilogramm schwere Kugel mit vier Antennen sauste um die Erde, bis sie am 4. Januar 1958 beim Wiedereintritt in die Erdatmosphäre verglühte. Als Reaktion auf den Sputnik starteten die USA ein insgesamt 25,4 Milliarden Dollar teures Forschungsprogramm, das die „Schmach des Sputniks“ (so Präsident John F. Kennedy) erst mit der Mondlandung von Neil Armstrong am 21. Juli 1969 ausgleichen konnte. Auf dem Weg dorthin wurde auch das Internet erfunden.

Gute Gelegenheit

Verantwortlich für den Wettlauf der Systeme war auf der amerikanischen Seite ARPA, die Advanced Research Projects Agency. Sie wurde mit dem Befehl 5105.15 des Verteidigungsministeriums am 7. Februar 1958 gegründet und nahm bereits im März ihre Arbeit auf. Die zeitliche Abfolge suggeriert, dass die ARPA in Reaktion auf den Sputnik entstand, doch dies darf man im Licht neuer historischer Forschungen als Mythos abtun. Der piepsende Satellit war für eine Gruppe von einflussreichen Planern um Vannevar Bush, James Conant und James Kilian nur eine hervorragende Gelegenheit, schneller die Forschungsagentur ins Leben zu rufen, die gegen die Sowjetunion antreten sollte.

Bereits 1950 hatten sich Bush, Covenant und Kilian besorgt darüber geäußert, dass die amerikanischen Studenten die falschen Fächer studierten. 7,8 Millionen heimgekehrte Weltkriegssoldaten machten von der sogenannten „G. I. Bill“ Gebrauch und studierten vor allem geisteswissenschaftliche und ökonomische Fächer. Mit harten Wissenschaften, gar mit Rüstungsforschung, wollten die Heimkehrer nichts zu tun haben. Um den Ernst der Entwicklung zu verdeutlichen, gab James Conant, damals Leiter des National Science Bureaus, eine Studie bei Nicholas De Witt von der Universität Harvard in Auftrag.

De Witt untersuchte 1954 das sowjetrussische Universitätssystem und veröffentlichte 1955 seine alarmierenden Ergebnisse: „Wir verlieren den neuen Krieg. Wir verlieren ihn, weil wir im Rennen verloren haben, mehr und bessere Ingenieure und Wissenschaftler zu produzieren als die Kommunisten.“ Mit Ergebnissen dieser Studie bewaffnet, leistete James Kilian, Präsident des Massachusetts Institute of Science (MIT) die Lobbyarbeit und sorgte als Wissenschaftsberater des Präsidenten Eisenhower dafür, dass eine Agentur für Spitzenforschung aufgebaut wurde. Nach dem Flug des ersten Sputniks nannte man Kilian zwar nur noch „Raketenzar“, doch hatte dieser die Pläne für den Aufbau der ARPA längst abgezeichnet, als das Piepsen begann.

Computerzar und Raketenzar

Mit gleichem Recht hätte man den gelernten Journalisten Kilian auch als „Computerzar“ bezeichnen können, denn unter seiner Ägide entwickelten Wissenschaftler am MIT das Prinzip des Timesharing zur besseren Ausnutzung von Rechnerkapazitäten. Etliche Millionen Dollar flossen in das von Kilian befürwortete Project MAC (Multi Access Computer), für das der Psychologe J. C. R. Licklider gewonnen wurde. Ein weiteres wichtiges, noch von Kilian angestoßenes Projekt war die Entwicklung eines Satelliten-Kommunikationsnetzwerkes. Als Leiter des Technological Capabilities Panel (TCP) der CIA (diese Position wurde geheim gehalten) war er frühzeitig mit Überlegungen beschäftigt, dass Ausspähen der Sowjetunion mittels der U2-Langstreckenflieger durch Satelliten zu ersetzen, die Daten in verschlüsselten Päckchen empfangen und senden sollten. Als größte Leistung Kilians wird heute bewertet, 1960 aus der ARPA die NASA als Langzeitprojekt heraus-

gelöst und über die ARPA kleinere, gezielt forschende Projekte verwaltet zu haben. So entstand beispielsweise das Projekt Vela, mit dem Ziel, Atombombentests auf der ganzen Welt zu messen.

Aus den Forschungen für das Satellitenkommunikationsnetz entwickelte der bereits erwähnte Psychologe Licklider seine Idee eines „Intergalactic Computer Network“, eines weltumspannenden Kommunikationsnetzes für die Forschungsrechner der ARPA. Bei den ersten Entwürfen beruhte dieses Netz, das als Weiterentwicklung der paketbasierten Satellitenkommunikation gesehen wurde, auf den Ideen zum Timesharing-System, wo ein Task auf dem Host die Steuerung der Kommunikation übernehmen sollte. Als das sogenannte ARPAnet schließlich an den Start ging, übernahmen kleinere dezidierte Kommunikationscomputer namens Interface Message Processors (IMP) diese Aufgabe.

Im Sommer 1968 schrieb die ARPA den Auftrag zur Entwicklung dieses Spezialcomputers aus. 140 Firmen wurden angeschrieben, darunter solche Riesen wie IBM, Univac und Honeywell, aber auch Newcomer wie DEC. Den Auftrag holte sich schließlich die kleine Bolt Beranek und Newman (BBN) in Cambridge/Massachusetts, eine Ausgründung von MIT-Forschern – das ARPAnet und später das Internet konnten entstehen. Im Kontext des Kalten Krieges ist die Entstehung dieses Computernetzwerkes ein Produkt des Forschungswettlaufes mit der Sowjetunion, ein Nebenprodukt von Atomtechnologie und Weltraumfahrt.

In einer der ersten Erwähnungen zum Aufbau des ARPAnets schrieb die Time, dass das Netz es Wissenschaftlern gestatten sollte, beim Ausbruch eines Atomkriegs in Kontakt zu bleiben. Offenbar hatte der Reporter sich von der Darstellung eines paketbasierten Netzwerkes inspirieren lassen, in dem Datenpäckchen über verschiedene IMP-Knoten zum Ziel gelangen. Weder in den Ausschreibungsunterlagen noch in den verschiedenen internen Memoranden zum Netz findet sich der Begriff Atomkrieg. Doch hartnäckig hält sich die Idee vom Netz, das Atomschläge überleben sollte, genau wie die Geschichte vom Sputnik und der Gründung der ARPA. (JS)

DETLEF BORCHERS

ist freier DV-Journalist und arbeitet für deutsche und amerikanische Fachzeitschriften.

Literatur

[1] Katie Hafner, Matthew Lyon; ARPA KadabraK; Die Geschichte des Internet; dpunkt-Verlag, 1997

[2] James Kilian; Sputnik, Scientists and Eisenhower; A memoir of the first special assistant to the President for science and technology; The MIT Press, 1978

[3] Julian C. Lucena; Defending the Nation; U.S. Policymaking to create Scientists and Engineers from Sputnik to the ‚War against Terrorism‘; University Press of America, 2005

[4] Mitchell Waldrop; The Dream Machine; J.C.R. Licklider and the Revolution that made Computing Personal; Penguin Putnam, 2001

[5] Donald E. Walzenbach; Origins of CIA Science and Technology Directorate; http://cryptome.org/cia-spy-tech/cia-spy-tech.htm

Juristische Aspekte von Produktsperren

Etappenweise

Tobias Haar

Häufig bauen Hersteller technische Sperren in ihre Produkte ein – meist, um dadurch ihre Kaufpreisansprüche abzusichern oder die Nutzung ihrer Produkte zu kontrollieren. Wer Softwaresperren verwendet, bewegt sich aber in einer juristischen Grauzone mit einigen Tücken.

Apple ist mit dem iPhone mal wieder ein kommerzieller Geniestreich gelungen. Das Gerät verkauft sich bestens und viele mehr wollen es haben. Neben Apple möchten aber auch die Mobilfunkgesellschaften von diesem Hype profitieren. Deswegen versehen sie das Gerät selbst bei Abschluss eines längerfristigen Vertrages mit einem SIM-Lock, das die Benutzung mit einer anderen SIM-Karte unterbindet.

Die Bindung an den Mobilfunkanbieter – der Apple einen gewissen Prozentsatz an den mit dem Gerät erzielten Umsätzen zahlen muss – ist damit perfekt. Bei der exklusiven Vereinbarung zwischen Apple und T-Mobile für die Vermarktung des iPhones in Deutschland gehen einige von einer Umsatzbeteiligung in Höhe von zehn Prozent aus. In Deutschland ist das Anbieten von SIM-Lock-gesperrten Mobiltelefonen rechtlich nicht zu beanstanden, solange die Kunden darauf hingewiesen werden und dies den Gegebenheiten des Marktes entspricht. Eine einstweilige Verfügung zwang T-Mobile erst zur Lockerung seiner Vertragsbedingungen für das iPhone, das daraufhin auch mit SIM-Karten anderer Anbieter funktionieren sollte. Doch das Hamburger Landgericht entschied, dass die exklusive Bindung an T-Mobile zulässig sei.

Kurz nach der Veröffentlichung des iPhone wurden in den USA die ersten Hackerangriffe auf die SIM-Locks veröffentlicht und die ersten SIM-Lock-freien Geräte bei eBay angeboten. In allen Fällen ging Apple mit juristischen Mitteln dagegen vor. Andererseits verklagten Nutzer das Unternehmen vor kalifornischen Gerichten wegen einer Verletzung des Wettbewerbsrechts durch Einbau der SIM-Sperre.

Andere Länder – andere Rechtslage

In Frankreich verlangt das Gesetz den Vertrieb SIM-Lock-freier Geräte. Diese werden dann zwar vermutlich zu höheren Preisen vertrieben, aber es gibt sie. Auch bei subventionierten Handys mit SIM-Lock verlangt das französische Recht, dass diese auf Wunsch des Kunden zu entsperren sind. Der Anbieter darf dafür nur in den ersten sechs Monaten ab Vertragsschluss eine Bearbeitungsgebühr verlangen.

Diese Beispiele verdeutlichen, mit welchen kommerziellen Interessen und rechtlichen Schwierigkeiten man bei der Durchsetzung seiner Interessen mit technischen Sperren rechnen muss – sei es im Massengeschäft oder im Bereich hochwertiger Investitionsgüter.

Mit softwaretechnischen Nutzungs- und Verwendungssperren lassen sich auch in anderen Bereichen Interessen durchsetzen. So gibt es Hersteller hochwertiger Produktionsanlagen, die ihre Produkte mit Softwaresperren oder schlüsselabhängigen Aktivierungsmechanismen versehen. Ist ein Käufer wegen der hohen Anschaffungskosten nicht in der Lage, den gesamten Kaufpreis auf einmal zu zahlen, kann er mit dem Hersteller Ratenzahlung vereinbaren.

Wenn aber die Anlage oder Maschine in ein Land geliefert werden soll, in dem es im Zweifel keinen wirklich effektiven Rechtsschutz zur Durchsetzung noch offener Forderungen – also beispielsweise der Ratenzahlungen oder alternativ der Rücknahme der noch nicht bezahlten Kaufsache – gibt, muss man sich eben andere Mittel zum Schutz seiner Ansprüche einfallen lassen. Der Verkäufer könnte die verkaufte Anlage mit einer zeitlich befristeten Softwareaktivierung versehen, die es dem Käufer erlaubt, die Anlage nur bis zur Fälligkeit der nächsten Rate in vollem Umfang zu nutzen. Bezahlt der Käufer pünktlich, erhält er einen Freischaltcode, nach dessen Eingabe die Anlage wieder eine Zeit lang läuft, bis bei der nächsten Fälligkeit wieder ein neuer Code erforderlich wird. Und so weiter.

Alternativ dazu kann man auch vorsehen, dass sich ohne Eingabe eines neuen Codes die Performance der Maschine verlangsamt und sie weniger produziert. Daneben bietet sich an, ohne Code bestimmte nützliche Features der Anlage abzuschalten, bis die Rate bezahlt ist. Je nach Art der verkauften Güter und Höhe der durchzusetzenden Forderungen sind weitere Spielarten dieser Methode zur Durchsetzung von Forderungen denkbar. Rechtlich sind sie jedoch alle bedenklich, wenn man bei der Vertragsgestaltung nicht aufpasst.

Nach dem Kaufvertragsrecht ist der Verkäufer verpflichtet, dem Käufer die Sache zu übergeben und ihm Eigentum daran zu verschaffen. So steht es in Paragraph 433 des Bürgerlichen Gesetzbuches. Weiter heißt es dort, dass die verkaufte Sache frei von Mängeln sein muss. Ein Mangel liegt dann vor, wenn der Käufer die Sache nicht so nutzen kann, wie es vertraglich vereinbart wurde. Man spricht unter Juristen in so einem Fall von einer für den Käufer ungünstigen Abweichung der Ist- von der Sollbeschaffenheit.

Kann ein Käufer eine Anlage nach Ablauf einer gewissen Zeit nicht mehr nutzen, ist diese Sache mangelhaft. Denn der Verkäufer hat im Vertrag ja eine dauerhaft nutzbare Anlage verspro-

chen. Erfüllt sie diese Voraussetzungen nicht und ist daran bei Einbau einer Softwaresperre der Verkäufer schuld, kann der Kunde ihn wegen Mängeln der verkauften Sache in Anspruch nehmen. Der Käufer kann Ersatz der mangelbehafteten durch eine mangelfreie Anlage verlangen, vom Kaufvertrag zurücktreten oder den Kaufpreis mindern sowie Schadensersatz verlangen – und das kann bei Produktionsanlagen, die längere Zeit nicht produzieren können, beliebig teuer werden.

Da ein Mangel aber nur vorliegt, wenn die Kaufsache nicht derjenigen entspricht, die im Kaufvertrag beschrieben wurde, liegt kein Mangel vor, wenn sich die Vertragsparteien auf die Vereinbarung etwa von Ratenzahlung und daran gekoppelte Softwaresperren verständigen. Steht im Kaufvertrag ausdrücklich, dass die gekaufte Anlage nur für eine bestimmte Zeit läuft, dann einen neuen Freischaltcode benötigt, den es nur gibt, wenn der Käufer die entsprechende Rate zahlt, ist juristisch alles in bester Ordnung.

Vorsicht AGB-Falle

Aber Vorsicht. Bei Standardverträgen ist auch das nicht so ohne Weiteres der Fall. Wenn der Verkäufer in seinen allgemeinen Geschäftsbedingungen – vielleicht sogar noch an etwas versteckter Stelle – vorsieht, dass er bei Nichtzahlung die Anlage außer Betrieb setzen kann, hilft ihm das nicht. Denn nach dem strengen deutschen AGB-Recht handelt es sich in einem solchen Fall um eine unwirksame, weil überraschende Klausel, auf die sich der Verkäufer nicht berufen kann. Steht die Anlage dann still, geht die Rechtsprechung konsequenterweise von einem Mangel aus, für den der Verkäufer wie beschrieben geradestehen muss. Nicht genug, dass er seine Kaufpreisforderung nicht wirksam durchsetzen konnte, es drohen ihm außerdem schmerzhafte finanzielle Forderungen des Käufers.

Juristische Auseinandersetzungen sind ebenfalls vorprogrammiert, wenn der Verkäufer den Freischaltcode zu spät oder gar einen falschen verschickt. Auch dann kann sich der Käufer beim Verkäufer für seine Schäden schadlos halten – jedenfalls soweit, wie die Haftung des Käufers nicht wirksam nach dem Kaufvertrag beschränkt wurde.

Eine wichtige Rolle spielt in solchen Fällen die Frage, welches Recht auf einen Kaufvertrag anzuwenden ist. Schließen zwei Unternehmen aus verschiedenen Ländern einen Kaufvertrag, sollten sie sich darin einigen, welche der beiden nationalen Rechtsordnungen für ihren Vertrag gelten soll. Andernfalls muss man auf der Basis der Regeln des sogenannten Internationalen Privatrechts klären, welche Rechtsordnung etwa bei Mängeln Anwendung findet. Eine weitere Besonderheit stellt das sogenannte UN-Kaufrecht dar. Bei ihm handelt es sich um ein „Spezial-Kaufrecht“, das für alle internationalen Kaufverträge gilt, wenn Käufer und Verkäufer ihren Sitz in Ländern haben, die jeweils das UN-Kaufrecht ratifiziert haben. Deutsches und UN-Kaufrecht unterscheiden sich aber kaum voneinander. Im Bereich der Softwaresperren sind die Ergebnisse der rechtlichen Bewertung meist gleich.

Kein rechtliches Problem entsteht, wenn der Hersteller einem Kunden im Rahmen eines Schnupperangebotes Software kostenlos für einen bestimmten Zeitraum überlässt und diese danach nur noch genutzt werden kann, wenn der Interessent einen entgeltpflichtiger Aktivierungscode erwirbt.

Unkomplizierter aus juristischer Sicht sind auch Maßnahmen, wie sie jetzt beim 100-Dollar-Laptop geplant sind. Dort sollen Geräte mittels ferngesteuerter Softwaresperren deaktiviert werden können, wenn diese in falsche Hände geraten, sprich geklaut werden. Bittet der Käufer den Verkäufer um Aktivierung einer solchen Sperre, kann selbstverständlich kein Mangel vorliegen. Dies wäre höchstens der Fall, wenn die Sperre trotz vertraglicher Vereinbarung in Wirklichkeit gar nicht funktioniert oder sie ein Dieb durch einfachste Mittel umgehen könnte. Verkauft jemand aber einen solchen Laptop legal weiter und die Sperre ist aktiviert, droht dem Weiterverkäufer, dass man ihn auf Gewährleistung in Anspruch nimmt.

Wenn eine Softwaresperre juristisch nicht zu beanstanden ist, darf ein Käufer sie auch nicht ohne Weiteres beseitigen. Apple geht mit technischen und juristischen Mitteln gegen die Hacker von iPhones vor. Updates enthalten eine Funktion, die das iPhone nach einem Hacking wieder mit einer Softwaresperre versieht. Nach Entfernen der Sperre aufgespielte Programme werden deaktiviert, die Geräte teilweise in den Auslieferungszustand zurückversetzt.

Wer den Quellcode der Steuerungssoftware auf den iPhones (oder in jedem anderen softwaregesteuerten Gerät) knackt und bearbeitet, verstößt damit gegen das Urheberrechtsgesetz. Neben Ansprüchen auf Unterlassung und Schadensersatz drohen in einigen Fällen auch Geld- und Freiheitsstrafen. In Einzelheiten bestehen hier durchaus Unterschiede zwischen den gesetzlichen Regelungen der verschiedenen Länder. Ist ein Produkt rechtswidrigerweise mit einer Softwaresperre versehen, muss der Käufer vom Verkäufer die Beseitigung der Sperre verlangen und notfalls einklagen. Eine eigenmächtige Beseitigung der Sperre im Wege der Selbstjustiz kommt nur in Not- und Ausnahmefällen in Betracht.

Neben rein kommerziellen Forderungen können auch andere Interessen mithilfe von Softwaresperre und Aktivierungszwang durchgesetzt werden. Wenn ein Softwarehersteller eine Registrierung seiner Produkte verlangt, kann er die installierte Software nach Ablauf einer gewissen Zeit einfach lahmlegen, wenn der Nutzer sie nicht nach erfolgter Registrierung mit einem Freischaltcode füttert. Oder ein Softwarehersteller ist an den persönlichen Daten der Nutzer interessiert. Neben den kaufrechtlichen Erwägungen kommt hier noch hinzu, dass es möglicherweise sittenwidrig sein kann, wenn ein Verkäufer die uneingeschränkte Nutzung seiner Produkte von der Bereitstellung entsprechender Daten abhängig macht. Diskutiert wurde das unter Juristen beispielsweise im Zusammenhang mit Werbeeinblendungen im Gegenzug für kostenlose Telefongespräche.

Fazit

Wer den Aufwand der Durchsetzung von Forderungen mithilfe staatlicher Gerichte und anderer Einrichtungen scheut oder Angst haben muss, dass in bestimmten Ländern Forderungen einfach nicht durchsetzbar sind, sollte über die Verwendung von Softwaresperren nachdenken. Bei der Gestaltung der entsprechenden Verträge ist aber Vorsicht angezeigt. Nur wenn die Softwaresperre klar beschrieben ist und auch entsprechend den Vereinbarungen genutzt wird, ist man rechtlich auf der sicheren Seite. In anderen Fällen drohen Schadensersatzforderungen und eine Verurteilung zur Beseitigung der Softwaresperre. (ur)

TOBIAS HAAR, LL.M.,

ist Rechtsanwalt mit Schwerpunkt IT-Recht.





Herstellerstrategien gegen Energieverschwendung

Grüner surfen

Barbara Lange

Rechenzentren und Netzinfrastruktur verbrauchen zu viel Energie. IT-Hersteller und Politik entwickeln dagegen Pläne, die sowohl das Klima als auch die Umsätze retten sollen.

Die Analysten von Gartner haben kürzlich „Grüne IT“ auf Platz eins der wichtigsten strategischen Technologien für 2008 gesetzt. Rechenzentrumsbetreiber sollten, so die Empfehlung der Marktforscher, möglichen gesetzlichen Regulierungen mit eigenen Strategien zuvorkommen. Denn zwischen 2000 und 2005 habe sich der Stromverbrauch von Servern in den USA verdoppelt. Die Informationstechnik zeichnet für zwei Prozent des globalen CO_2-Ausstoßes verantwortlich und liegt damit auf dem Niveau des internationalen Flugverkehrs, schätzt Gartner.

Effizientere Prozesse mit IT

Zwar ist die Rolle von IT und Internet zweischneidig: Zum einen Stromfresser und CO_2-Verursacher, zum anderen Lösungslieferant für Unternehmen, die ihre Prozesse mithilfe von IT umweltverträglicher und effizienter gestalten können. Dennoch ist Konsens, dass die IT-Branche einen Beitrag zur Entlastung der Umwelt leisten muss, und auch die höchste politische Ebene hat die Relevanz erkannt. So organisierte das Bundesumweltministerium gemeinsam mit dem Institut für Zukunftsstudien und Technologiebewertung (IZT) den Workshop „Grüner Surfen – Perspektiven für eine energieeffiziente Nutzung des Internets“ mit dem Ziel, Provider, Hersteller, politische Verantwortliche und Nutzer zunächst einmal für dieses Thema zu sensibilisieren und Erfahrungen auszutauschen.

Als Referenten traten Vertreter von Sun Microsystems, T-Com, Cisco, Strato, Igel, Nine Internet Solutions, Bitkom und Verbraucherverbänden auf.

Dr. Ulf Jaeckel, Leiter des Referats „Produktbezogener Umweltschutz, Normung“ im BMU, gab den 45 Teilnehmern aus Politik, Wirtschaft und Wissenschaft einen Überblick über die Pläne der Bundesregierung. In den nächsten Jahren will man intelligente Messverfahren entwickeln, mit denen die Hersteller ihre Produkte kennzeichnen. Perspektivisch sollen Bund, Länder und Gemeinden nur noch energieeffiziente Produkte kaufen. Hierfür erarbeitet man derzeit gemeinsam mit dem Branchenverband Bitkom einen Leitfaden. Denkbare Maßnahmen wären auch ein geringerer Mehrwertsteuersatz für sparsame Geräte.

Als Haupt-Stromverbraucher gelten die Rechenzentren, denn Server, Netzwerk und Kühlung brauchen viel Energie, arbeiten aber oft ineffizient. Eines der vielen Probleme dabei ist die Auslastung der Infrastruktur, die häufig nur bei 15 Prozent liegt.

Rechenzentren mit schlechter Auslastung

Allein durch den Einsatz vorhandener Technologien könnte man den Stromverbrauch halbieren, rechnete Jaeckel vor. Den Versuch, die Watt-, CO_2- und Euro-Rechnerei vorstellbar zu gestalten, unternahm Rolf Kersten von Sun Microsystems. Nach seinen Berechnungen entstehen für eine Kilowattstunde Strom in Deutschland 600 Gramm Kohlendioxid.

Auf der Basis dieses Verhältnisses braucht eine Google-Abfrage vier Wattstunden und erzeugt drei Gramm CO_2 pro Suchabfrage. Gar nicht mal so schlecht, denn die Infrastruktur von Google sei mit einer Auslastung von 93 Prozent gut optimiert, im Gegensatz etwa zu den Servern von Linden Labs. Ein Second-Life-Avatar braucht in einem Jahr – wenn er immer online ist – 195 kWh und atmet so viel CO_2 aus wie vier „richtige“ Menschen im gleichen Zeitraum. Der Stromverbrauch der PCs der Nutzer ist dabei nicht mit eingerechnet.

Lösungsansatz Virtualisierung

IT-Hersteller haben dies erkannt. Cisco, mit dem kürzlich eröffneten Themenportal www.gruene-it.org engagiert in Sachen „Grüne IT“, setzt auf Virtualisierung – ein Ansatz, der den Ausnutzungsgrad der Speicherkapazität von derzeit 30 auf bis zu 70 Prozent erhöhen kann. Beim Sparen helfen können auch Videokonferenzen, die die Reisetätigkeit eindämmen und die Arbeit virtueller Teams unterstützen. Cisco unterhält derzeit 130 Telepräsenz-Installationen. Damit habe das Unternehmen seinen eigenen CO_2-Ausstoß um 10 Prozent verringern und die Reisekosten messbar senken können, so der Vertreter des Netzwerkausrüsters.

Inwieweit es den Unternehmensleitungen bei ihren Maßnahmen um die Rettung der Erde geht, dürfte sich kaum feststellen lassen. Dass klimaschonendes Engagement als wichtiger Baustein des Image-Marketings angesehen wird, ist aber nahezu Konsens. Und auf jeden Fall wollen und müssen

Firmen ihre Energiekosten senken, um wettbewerbsfähig zu bleiben.

Zum Beispiel Strato. Investitionsentscheidungen werden dort seit zwei Jahren immer stärker durch die Frage der Energieeffizienz bestimmt, so Damian Schmidt, Vorstandsvorsitzender des Unternehmens, in seinem Vortrag. Wo 30 000 Server in zwei Rechenzentren täglich über 200 Millionen E-Mails empfangen und weiterleiten sowie 3,5 Millionen Domains hosten, sei Energie der höchste Kostenfaktor im RZ-Betrieb.

In den vergangenen 18 Monaten konnte das Unternehmen durch eine Optimierung in den Bereichen Hard- und Software sowie Gebäudetechnik 30 Prozent Energie sparen (*iX* berichtete). Und: Ab 2008 will Strato zu 100 Prozent Energie aus Laufwasserkraft nutzen.

Nebenbei bemerkt: Wieviel CO_2 durch das Gesetz zur Vorratsdatenspeicherung erzeugt werden wird, lässt sich noch nicht hochrechnen. Cisco zum Beispiel schätzt, dass nur vergleichsweise wenig Kapazitäten zusätzlich aufgebaut werden müssen.

Abhängig wird dies unter anderem von der Wahl der Speichermedien sein (Band/Festplatte), die wiederum davon abhängt, wie die Behörden auf die bevorrateten Daten zugreifen wollen.

Zurzeit entsteht ein Kostendruck durch die Anschaffung der Hardware: Allein die Technik wird bis zu 75 Millionen Euro kosten, schätzt Bitkom.

Moderner Ablasshandel in der Kritik

Einen anderen Ansatz verfolgt die Deutsche Telekom. Der Stromverbrauch des Unternehmens steigt, und durch eine laufende Umstellung der Netze auf Internet-Technik gibt es wohl derzeit keine andere Chance, als Energieverbrauch und CO_2-Emission zu entkoppeln, so Claudia Schwab, Leiterin Umweltschutz und Nachhaltige Entwicklung bei T-Com. Daher verfolgt das Unternehmen seit drei Jahren das Prinzip der Klimakompensation: Man berechnet den eigenen CO_2-Ausstoß und finanziert ein Klimaschutzprojekt, meistens in Entwicklungsländern, das die eigene Emission ausgleicht.

Diese von Kritikern als moderner Ablasshandel bezeichnete Methode der Klimaneutralisierung nutzt auch das Schweizer Unternehmen Nine Internet Solutions, das 65 Tonnen CO_2 im Jahr allein durch sein Rechenzentrum produziert. Zunächst versuchte man, den nach dem Einbau einer neuen Servergeneration explodierten Stromverbrauch durch Virtualisierung und Free-Cooling-Kühlsysteme zu optimieren, dann aber war nichts mehr drin. Durch eine Kooperation mit dem Unternehmen Myclimate und eine Investition in Klimaschutzprojekte kompensieren sie ihren CO_2-Ausstoß.



Wer Videokonferenzen nutzt, muss nicht fliegen oder fahren: Cisco setzt auf TelePresence.

Mehr als Stromsparen

Den Energiebedarf senken können auch Thin Clients. Nach einer Studie des Fraunhofer Institut für Umwelt-, Sicherheits- und Energietechnik (UMSICHT) verbrauchen sie mit 40 Watt, inklusive Serveranteil und Kühlung, nur halb so viel Energie wie voll ausgestattete PCs.

Grüne IT ist mehr als Strom sparen. Noch zu oft werden die ökologischen Belastungen bei der Produktion und Entsorgung von Hightech-Geräten außer Acht gelassen. Mit dem Recyceln von Elektroschrott in Ländern wie Indien, China und Südafrika beschäftigte sich unter anderem die Konferenz „Sustainable IT“, veranstaltet von Newthinking store GmbH und der Amina-Stiftung. Nach einem Vortrag von Martin Streicher, Forschungsinstitut Empa in St. Gallen, belastet besonders in Indien und China ein „Hinterhofrecycling“ Mensch und Umwelt. Elektroteile werden schlicht auf der Straße oder in Hinterhöfen ohne irgendwelche Schutzmaßnahmen auseinandergelötet und sortiert, so Streicher. In Südafrika sei das Recycling vergleichsweise am stärksten als offizielles Business etabliert.

Einen Schwerpunkt wird die Grüne IT auch auf der Cebit 2008 haben. Angekündigt sind ein spezielles Kongressprogramm und eine „Green-IT-Village“, in der spezielle Lösungsansätze gebündelt sind, zum Beispiel ein „grünes Rechenzentrum“. (JS)

BARBARA LANGE

ist IT-Journalistin und Inhaberin des Redaktionsbüros kurz&einfach in Lengede.

Onlinequellen

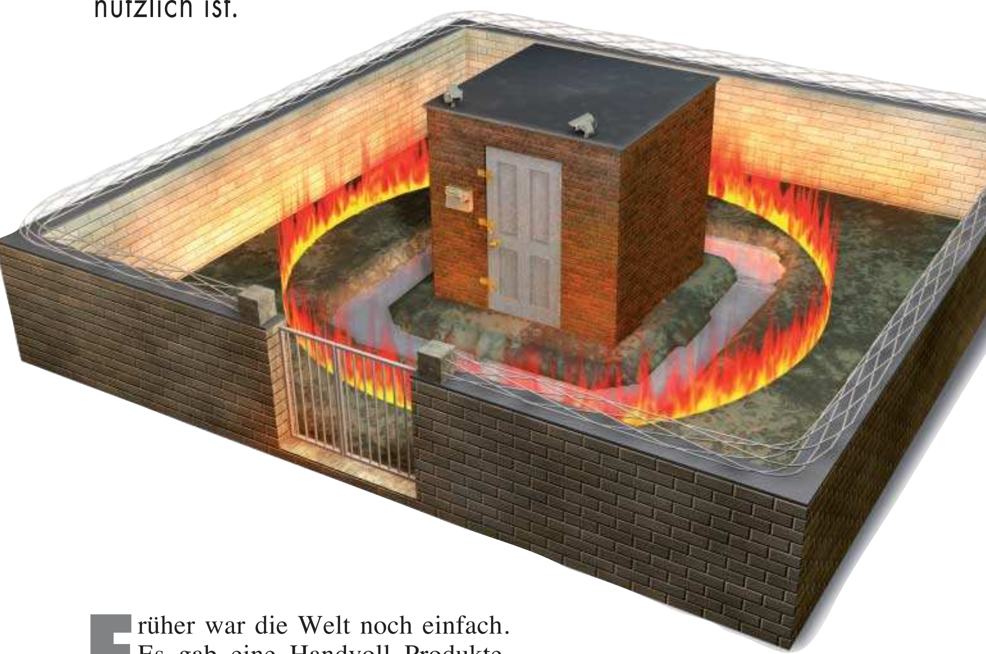
Blog von Rolf Kersten, Sun Microsystems	http://blogs.sun.com/rolfk
Konferenz „Sustainable IT“	www.sustainable-it.org
Themenportal „Grüne IT“ von Cisco	www.gruene-it.org
Grüne IT als Top-Thema bei Gartner	www.gartner.de/fokus/071042_10.html
Thin-Client-Studien vom Fraunhofer Institut UMSICHT	http://it.umsicht.fraunhofer.de/TCecology/ http://it.umsicht.fraunhofer.de/PCvsTC/

Königsweg Unified Threat Management?

Multischutz

Jörn Maier

Mehr und mehr Features haben Hersteller in den letzten Jahren in ihre Sicherheitsprodukte integriert. Wer sich eine solche UTM-Lösung zulegt, sollte sich zunächst mit ihnen vertraut machen und entscheiden, was fürs eigene Unternehmen nützlich ist.



Früher war die Welt noch einfach. Es gab eine Handvoll Produkte, mit deren Hilfe ein Netzwerk geschützt werden konnte und musste. Eine Firewall zur Verteidigung gegen Angriffe aus dem Internet, einen zentralen Virenschutz-PC, an dem die Mitarbeiter ihre Datenträger prüfen lassen sollten (was sie meist sowieso nicht taten), damit keine Schädlinge ins hauseigene Netz gelangen, und wer sehr viel Geld in seinem Budget hatte, durfte sich auch ein Intrusion-Detection-System (Intrusion-Prevention-Systeme gab es erst viel später) anschaffen. Und natürlich stand über allem das Paradigma der unterschiedlichen Hersteller. Schließlich wollte man so das Risiko von Systemfehlern und damit potenziellen Schwachstellen streuen.

Doch diese Zeiten sind vorbei. Die Anforderungen an Sicherheitssysteme sind komplexer und deren Administration zeitaufwendiger geworden. Eine einzelne Firewall existiert in kaum einem großen Unternehmen mehr, meist sind es viele. Der zentrale Virenschutz-PC wurde durch ein umfassende Anti-Virus-Enterprise-Management-Lösung ersetzt und die Konfiguration weiterer Dienste wie VPN, URL- und Content Filtering, Network Access Control (NAC) und andere sind schon lange in das Aufgabengebiet der Sicherheits- oder Systemadministratoren gewandert.

Eine bestimmte Gattung an Sicherheitssystemen hat sich unter dem Schlagwort „Unified Threat Management" (UTM) in den letzten Jahren besonders hervorgetan, meist in Form von sogenannten Appliances, einer Kombination von Hard- und Software. Es gibt in der Zwischenzeit keinen ernstzunehmenden Hersteller von Sicherheitsprodukten mehr, der nicht ein solches Gerät – meist eine gesamte Familie – in seiner Angebotspalette hat. Und die Anforderungen an diese Produkte und ihre Aufgaben steigen.

Für und gegen alles

Den Begriff UTM definierte im Jahr 2004 Charles Kolodgy von IDC. Er beschrieb eine neue Gattung von Produkten, die Firewall, IDS/IPS und Anti-Virus in einem Gerät vereinen. Ursprünglich war diese Gerätegattung auf kleine und mittelständische Unternehmen abgestimmt, in denen es keine spezialisierten Sicherheitsadministratoren gibt. Die durch die Hersteller propagierten Vorteile wie einfache Handhabung, geringer Platzverbrauch, günstiges Preis-Leistungs-Verhältnis et cetera wurden bald durch den „Plug-and-play-Mythos" ersetzt.

So versuchte manch ein Verkäufer seinen Kunden weiszumachen, dass der Kauf eines UTM-Gerätes ausreiche, das gesamte Netzwerk abzusichern. Die Administratoren würden nur minimal belastet und die Sicherheit drastisch erhöht. Auch die großen Unternehmen begannen sich für die „Wunderkisten" zu interessieren. Schließlich haben die meisten mehrere Außenstellen, die es abzusichern gilt. Dies führte zu weiteren Anforderungen an die UTM-Produkte. Dazu gehören vor allem Interfaces für das zentrale Management mehrerer solcher Geräte oder zumindest die Integration einzelner Dienste wie Virenschutz in die bestehende Enterprise-Management-Plattform.

Darüber hinaus veränderte sich der Personenkreis, der diese Produkte betreute. Waren es zuvor meist Administratoren, die kaum Zeit hatten, sich in neue Werkzeuge einzuarbeiten, kamen nun Security-Verantwortliche hinzu, die den Umgang mit komplexen Sicherheits-Tools gewohnt waren und auf deren Features nicht verzichten wollten.

Weitere Dienste wie Virtual Private Networks (VPNs) zur Anbindung von Außenstellen fanden Eingang in die UTM-Produkte. Darüber hinaus versuchten manche Hersteller, die Sicherung komplexerer Dienste wie Voice over IP (VoIP) durch Quality-of-Service-Mechanismen zu ermöglichen oder weitere Dienste wie URL- und Content-Filter-Funktionen sowie die Absicherung von Social-Networking-Plattformen in die Systeme zu integrieren. Und die

Entwicklung bleibt nicht stehen: Netzwerk-Optimierung, Data Leakage Protection, Session Border Controller et cetera sind nur einige der Schlagworte, die man früher oder später in irgendeiner Form in diesen Produkten umgesetzt finden wird und die vielleicht charakteristisch für die nächste Generation „UTM 2.0“ sein werden. Die alte Definition von UTM ist demnach schon lange passé.

Da es keine verbindliche Anzahl von Diensten gibt, die ein modernes UTM-Produkt ausmachen, sollte der Kunde zunächst die eigene Situation im Unternehmen analysieren. Dazu gehören Fragen, welche durch ein UTM-Gerät angebotenen Sicherheitsdienste wie Firewall, IDS/IPS, Anti-Virus, VPN, URL- oder Content Filter, Content Scanning, Spam-Filter et cetera das Unternehmen wirklich benötigt. Ein Betrieb, der eine Appliance lediglich zur zentralen Sicherung des Netzwerks betreiben möchte, hat andere Anforderungen als einer, der die Anbindung seiner Außenstellen mit einem zentralen Internetzugang über die Hauptgeschäftsstelle sicherstellen will.

Wird ein System im internen Netz eingesetzt, so kann es beispielsweise notwendig sein, dass das Gerät mit virtuellen LANs (VLANs) umgehen kann. Ein Intrusion Prevention System (IPS), das neben den normalen IP-Adressen und Ports auch VLAN-IDs versteht, kann in manchen Situationen deutlich besser agieren als ein Standardprodukt. Ein Großteil der UTM-Hersteller (siehe Hersteller-Tabellen) hat bereits erkannt, wie wichtig VLANs in der Praxis sind, und seine Produkte entsprechend ausgelegt.

Aber nicht nur VLANs, sondern auch die Integration in das normale Routing eines Netzwerks kann manchmal von Bedeutung sein. Zwar versucht man Security Appliances immer noch so weit es geht mit statischen Routen zu betreiben, es gibt jedoch komplexe Situationen, in denen das schlicht nicht mehr möglich ist. Dann sollte die UTM-Appliance in der Lage sein, die gewünschten Routing-Protokolle zu verstehen und sich sicher in das Netzwerk zu integrieren. Auch hier haben viele Hersteller ihre UTM-Geräte dahingehend erweitert.

Das zentrale Feature der UTM-Devices ist nach wie vor die Firewall. Fast alle Systeme arbeiten im „Stateful inspection“-Modus. Diesen allgemeinen Ansatz erweitern jedoch manche Hersteller um verschiedene Proxies für Protokolle jenseits von HTTP, SMTP, FTP. Genauer gesagt soll das UTM-Produkt überprüfen, ob auch wirklich das gewünschte Protokoll übertragen wird und nicht ein anderes auf dem gleichen Port. Proxies bieten die verschiedenen Produkte etwa für SIP, H323, NetBIOS, VNC, RTP, DNS und so weiter. Jedem sollte allerdings klar sein, dass der Einsatz dieser Proxies zu Lasten der Performance geht. Ein vollständiges Aktivieren aller Protokollprüfungen ist daher nicht immer sinnvoll. Ähnlich wie bei der Layer-7-Prüfung sollte man sich fragen, in welchen Fällen eine Virenprüfung von Datenströmen, die nicht per SMTP, FTP oder HTTP übertragen werden, sinnvoll ist. Diese Auswahlmöglichkeit anderer Protokolle unterstützt allerdings nicht jedes UTM-System.

Generell sollten die Verantwortlichen bei der Auswahl einer UTM-Appliance nicht nur die aktuelle Situation, sondern auch zukünftige Projekte oder Unternehmensstrategien in Betracht ziehen. Plant ein Unternehmen etwa, die Telekommunikation eines Unternehmens auf VoIP umzustellen, so sollte das Produkt weitere Features wie Quality of Service oder die bereits genannten Möglichkeiten zur Untersuchung des VoIP-Datenstromes via Firewall (eigener Proxy) bieten.

Darüber hinaus schadet es nicht, wenn man einen genaueren Blick auf die einzelnen Dienste innerhalb eines UTM-Device wirft. Die wenigsten Hersteller entwickeln alle Dienste selbst. Meist gehen sie Partnerschaften ein oder greifen gelegentlich auch zu Open-Source-Lösungen. Insbesondere beim Virenschutz sollten potenzielle Käufer darauf achten, dass Engine und Pattern von einem anderen Hersteller stammen als von demjenigen, dessen Produkt man bereits intern einsetzt. Ein Blick „unter die Haube“ stellt auch sicher, dass Erfahrungen, die man mit Einzelprodukten bereits gemacht hat (URL-, Content Filter, IDS/IPS et cetera), in die eigene Bewertung einfließen.

Eine Frage, die man sich vor dem Erwerb eines UTM-Gerätes stellen sollte, betrifft die Verfügbarkeit: Reicht ein einzelnes System aus oder ist die darüber laufende Kommunikation so unternehmenskritisch, dass ein Failover oder gar Clustering zu den zwingenden Anforderungen gehört? Welche Dienste übertragen die Daten? Ist es notwendig, dass bei einem Ausfall eines Systems die bereits aufgebaute Verbindung übernommen wird (Stateful Failover) oder reicht eine Neuinitiierung der Verbindung aus? Ist es notwendig, dass beide Systeme aktiv sind (aktiv/aktiv) oder kann ein System auch als Standby-System betrieben werden (aktiv/passiv)?

Nachrüsten von Vorteil

Die meisten Hersteller ermöglichen ein aktives Failover und immerhin zehn der von *iX* befragten geben an, ihre Systeme in einem Cluster betreiben zu können und dadurch zumindest eine gewisse Form von Load Sharing durchzuführen. Allerdings variiert die mögliche Anzahl der zu clusternden Geräte von 2 bis 10 (Astaro) stark. Generell sollte man auf Erweiterungsmöglichkeiten achten. Vor allem bei der geplanten Aktivierung rechenintensiver Dienste wie VPN oder Anti-Virus drohen sonst Performance-Einbrüche.

Und das Management wird einen Antrag auf Erneuerung eines ein Jahr alten Systems kaum begrüßen – auch wenn vorher nicht abzusehen war, dass nun plötzlich ein Service X über die UTM-Appliances laufen muss. Einige Geräte (Secure Computing, Securepoint, underground_8) lassen sich mittels Kryptobeschleunigern oder Koprozessoren ausbauen, was unter Umständen einiges an Geld sparen kann.

Generell ist die Performance einer der wichtigsten Punkte bei der Auswahl eines Produkts. Man sollte sich

iX-TRACT

- Seit ihren Anfängen haben sich UTM-Lösungen vom relativ einfach gestrickten 3-Feature-Modell zu komplexen Suiten mit zahlreichen Sicherheits- und Netzwerkfunktionen entwickelt.
- War UTM in den ersten Jahren vor allem für kleine und mittlere Unternehmen ohne eigenen Sicherheitsadministrator interessant, so ist es heute nicht nur dank hoher Performance auch für große Firmen von Nutzen.
- Vor der Anschaffung einer UTM-Lösung empfiehlt sich ein genauer Blick „unter die Haube“. Die Unterschiede zwischen den Produkten sind zwar nicht riesig, aber unter Umständen entscheidend.

Anbieter von Unified-Threat-Management-Lösungen

Allgemeine Angaben						
Hersteller	**Astaro**	**Checkpoint**	**Clavister**	**Collax**	**Fortinet**	**Funkwerk**
Webadresse	www.astaro.com	www.checkpoint.com	www.clavister.de	www.collax.com	www.fortinet.com	www.funkwerk-ec.com
Produkt	ASG 110; 120; 220; 320; 425a; 525	UTM-1; VPN-1 UTM; VPN-1 UTM Edge	SG10; SG50; SG3100; SG4200; SG4400	Security Gateway	FortiGate (versch.)	Funkwerk UTM 1100; 1500; 2100; 2500
Art des Produkts	Hardware-Appliance, Software	Hardware-Appliance, Software	Hardware-Appliance, Software	Hardware-Appliance, Software	Hardware-Appliance	Hardware-Appliance
Basisbetriebssystem	eigenes Linux	Check Point SecurePlatform (linuxbasiert), IPSO, Solaris	Clavister Firewall CorePlus	Pynix (linuxbasiert)	FortiOS 3.0	Linux
Features						
zusätzliche Services zu FW, IDS/IPS, Antivirus	VPN, URL-Filter, Antispam, Mail-Verschlüsselung u. a.	VPN, URL-Filter, Antispam, Web Application Firewall	VPN, Antispam, Content-Filter, Application Control, Loadbalancing u. a.	VPN, URL-Filter, Content-Filter, Antispam u. a.	VPN, Content-Filter, Antispam u. a.	VPN, URL-, Content-Filter, Antispam
Integration bestehender Produkte	✓	✓	–	von Kaspersky, IBM ISS (Cobion), NCP	–	✓
QoS (Zusich. v. Bandbreite)	✓	✓	✓	✓	✓	✓
Failover	aktiv/passiv, aktiv/aktiv	✓	aktiv/passiv, aktiv/aktiv, eig. Interface	link failover	aktiv/passiv, via FGCP	✓
Clustering	✓ bis zu 10 Einheiten	✓	k. A.	–	bis zu 4 Geräte	–
VLAN-Unterstützung	✓	✓	✓	✓	✓	–
Erweiterungsmöglichkeiten	nur SW-Version	nur SW-Version	✓	✓	3000er-Serie: AMC-Module via Interface	–
Layer 7 Inspection/ für welche Protokolle und Dienste	SMTP, POP3, HTTP, FTP, DNS, SOCKS, SIP	✓	SMTP, POP3, HTTP, FTP, TFTP, H.323, SIP	–	alle Protokolle per IPS, AV-Transparentproxies für SMTP, POP3, HTTP, FTP, IMAP, P2P, NNTP	SMTP, POP3, HTTP, FTP, DNS
unterstützte Routing-Protokolle	OSPFv2	RIP, RIPv2, OSPF, BGP, PIM	OSPF	–	RIP, RIPv2, OSPF, BGP, PIM	OSPF
Performance						
Anzahl der Interfaces	je nach Gerät 3 – 10	4 und 8	3 – 14	je nach Gerät	je nach Gerät	4; 4; 6; 6
Beschränkungen IP-Adressen	je Lizenz	–	–	–	–	lizenzabhängig
max. Anzahl Verbindungen pro Sek., alle Dienste aktiv	k. A.	k. A.	k. A.	keine Limitierung	je nach Gerätefamilie	k. A.
max. Anz. paralleler Verbindungen, alle Dienste aktiv	je nach Gerät 60 000 – über 1 000 000 (nur FW)	je nach HW und RAM	4000 – 5 000 000	keine Limitierung	je nach Gerätefamilie	je nach Gerät 20 000 – 500 000
max. Durchsatz, alle Dienste aktiv	je nach Gerät 100 MBit/s – 3 GBit/s (FW) 25 – 750 MBit/s (FW+IPS) 30 MBit/s – 400 MBit/s (VPN)	400 MBit/s (VPN) 2 GBit/s (FW)	50 – 4000 MBit/s	abhängig von HW	je nach Gerätefamilie	je nach Gerät 115 – 1900 MBit (FW) 25 – 540 MBit (FW+IDS) 20 – 210 MBit/s (VPN)
Management						
Manag. über zentr. Interface	✓	✓	✓	–	✓	–
Authentifizierungsmöglichkeiten	LDAP, RADIUS, AD, TACACS+	LDAP, RADIUS, AD, Zertifikate u. a.	LDAP, RADIUS, AD	LDAP, AD, PDC, Kerberos	LDAP, RADIUS, AD, PKI, TACACS, lokal	LDAP, RADIUS, lokal, OOBA
Benutzer- u. Rollenkonzept	in kommenden Versionen	✓	✓	✓	✓	–
Dienste unabhäng. auf versch. Interfaces einsetzbar	✓	✓	✓	✓	✓	✓
Reihenfolge der Dienste definierbar	–	–	–	–	–	–
Logging/Alarming/Reporting						
wie erfolgt Logging/ erlaubte Dateigröße	Syslog	zentraler Management-Server/ext. Syslog-Server	Syslog oder proprietär	Syslog/bei Appliance 150 GByte	Syslog oder proprietär	Syslog
welche Alarmmöglichkeiten	SNMP, E-Mail	SNMP, E-Mail, Alert via Popup	SNMP, E-Mail, SMS	SNMP, E-Mail	SNMP, E-Mail, Fortianalyzer	SNMP, SMTP
Log-Korrelation	✓	mit Eventia Analyzer	✓	mit Ereignis-Monitor	mit Fortianalyzer	–
Generieren von Reports möglich	Report Manager	Expressreports in manchen Prod. mögl.	✓	✓	mit Fortianalyzer bis zu 300 Reports	–
Sonstiges						
Preis (netto)	je nach Gerät 1635 – 45 935 €	ab 7500 $; ab 3500 $; ab 600 $	495 – 57 750 €	SW ab 295 €, Appliance ab 895 €	k. A., je nach Gerät	799 €; 1099 €; 2499 €; 5999 € (Gerät für 10; 25; 100; 250 Benutzer)
Anmerkungen	–	–	–	spez. für kleine mittelständ. Unternehmen	Lizenz unlimitiert, Beschränkung durch HW	Setup-Wizard, unlimitierte Zahl VPN-Tunnels (Beschränkung durch Hardware)

Tabelle beruht auf Herstellerangaben
✓ ja/zutreffend/vorhanden
– nein/nicht zutreffend/nicht vorhanden
k. A. keine Angabe

Gateprotect	IBM Internet Security Systems	Juniper Networks	NetASQ	Phion	Secure Computing	Securepoint
www.gateprotect.de	www.iss.net	www.juniper.net	www.netasq.com	www.phion.com	www.securecomputing.com	www.securepoint.de
GPO 125; GPA 400; GPX 800	IBM Proventia MX 1004; MX 3006; MX 5010	Secure Services Gateway SSG 5 bis SSG 550	F 50	netfence M3; netfence gateway nf-850	Sidewinder G2	Piranja; RC100; RC200; RC300; RC400
Hardware-Appliance	Hardware-Appliance	Hardware-Appliance	Hardware-Appliance	Hardware-Appliance, Software	Hardware-Appliance	Hardware-Appliance, Software
Debian 3.1/Admin-Client: Windows XP, Vista	Red Hat Linux	ScreenOS	eigenes	eigenes (linuxbasiert)	Secure OS	Securepoint Security OS
VPN, URL-Filter, Content-Filter, Antispam	VPN, Antispam, Web-Filter	VPN, URL-Filter, Antispam u. a.	–	VPN, URL-Filter, Antispam u. a.	VPN, Content-Filter, Antispam u. a.	VPN, URL-Filter, Content-Filter, Antispam
von IBM ISS (Cobion), Kaspersky, Commtouch, Open Source	k. A.	von Kaspersky, Surfcontrol, Symantec	von Kapsersky	von IBM ISS, Avira, Microdasys	von McAfee, Sophos	–
✓	–	✓	✓	✓	ab Q1 08	✓
✓	aktiv/passiv	aktiv/passiv, aktiv/aktiv, Sub-Second Stateful	✓	aktiv/aktiv	aktiv/passiv, aktiv/aktiv	✓
k. A.	–	✓	–	bis zu 2 Geräte	bis zu 5 Geräte	–
✓	–	✓	✓	✓	✓	–
✓	–	WAN-/LAN-Interfaces, Switching-Module	–	✓	CPU, RAM, RAID	Prozessoren, RAM, Interfaces
SMTP, POP3, HTTP, FTP, VNC, SIP, RTP	SMTP, POP3, HTTP, FTP	SMTP, POP3, HTTP, FTP, IMAP, DNS, NetBIOS, P2P, IM	✓	SMTP, POP3, HTTP, FTP, H323, SIP	applikationsspezifische Proxies	SMTP, POP3, HTTP, FTP, SIP, VNC, RTP
k. A.	OSPF	RIPv2, OSPF, BGP	RIP, RIPv2, OSPF	OSPF, RIP	RIP V2, OSPF, BGP	–
3; 6; 10	4; 6; 10	k. A.	3	4 – 12	4 – 26	je nach Gerät 3 – 16
–	je nach Gerät	–	–	–	✓, größere Geräte unbeschränkt	–
17 000; 28 000; 43 500	2125; 4100; 12 500	je nach Gerätefamilie 2800 – 15 000	k. A.	5000; 20 000	600 Sessions pro Sekunde und 400 Mbps	je nach Gerät, RC300: 19 000
250 000; 500 000; 1 000 000	101 000; 120 000; 150 000	je nach Gerätefamilie 4000 – 128 000	15 000	400 000; 800 000	2 Millionen	je nach Gerät, RC300: 1 048 000
je nach Gerät 250 MBit/s – 3GBit/s (FW) 45 MBit/s – 200 MBit/s (VPN)	je nach Gerät 100 – 1600 MBit/s (FW) 100 – 800 MBit/s (FW + IDS)	je nach Gerätefamilie 160 MBit/s – 1 GBit/s (FW) 40 – 500 MBit/s (VPN)	100 MBit/s	je nach Gerät 390 – 4009 MBit/s (FW) 300 – 850 MBit (FW + IDS)	3,2 GBit/s	je nach Gerät RC300: 63 – 98 MBit/s
✓	✓	✓	k. A.	✓	✓	✓
LDAP, RADIUS, AD, lokal	LDAP, RADIUS, Secure ID, Xauth	LDAP, RADIUS, Secure ID	LDAP, RADIUS, lokal	LDAP, RADIUS, AD, Secure ID, MSNT	LDAP, RADIUS, AD	LDAP, RADIUS und proprietär
✓	eingeschränkt	✓	–	✓	✓	✓
✓	✓	✓	✓	✓	✓	✓
–	–	–	–	eingeschränkt	✓	–
Syslog oder proprietär/ 3 GByte	Syslog oder proprietär	Syslog oder proprietär	Syslog	lokal, neuere Versionen ohne Größenbeschränk.	Syslog, lokal/definierbar	Syslog, externer Server
Mail	SNMP, SMTP	Syslog, SNMP, E-Mail	SNMP, E-Mail, Syslog	SNMP, E-Mail	SNMP, E-Mail, Pager	Mail, Pop-up-Fenster
–	mit SiteProtector Management	✓	–	✓	✓	frei konfigurierbar
✓	mit SiteProtector	mit Report Manager selbst definiert und andere	Reporting-Suite	✓	Security Reporter	✓
995 €; 1995 €; 2595 €	1546 €; 5490 €; 14 890 €	ab 700 $	822 €	je nach Gerät ab 5999 €; ab 19 280 €	ab 1500 €	495 €; ab 1030 €; ab 1570 €; ab 3220 €; ab 7020 €
Leasing möglich	verhaltensbasierter Virenschutz, virtual Patches	–	–	–	–	–

Anbieter von Unified-Threat-Management-Lösungen (Fortsetzung)

Allgemeine Angaben							
Hersteller	**Smoothwall**	**SonicWall**	**Symantec**	**underground_8**	**Vasco**	**Watchguard**	**ZyXEL**
Webadresse	www.smoothwall.net	www.sonicwall.com	www.symantec.com	www.underground8.com	www.vasco.com	www.watchguard.de	www.zyxel.de
Produkt	SmoothGuard 1000 UTM	NSA E7500; PRO 3060	Endpoint Protection 11.0	Limes MF	aXs Guard	Firebox X Edge; Core; Peak	ZyWALL USG 300
Art des Produkts	Hardware-Appliance	Hardware-Appliance	Software	Hardware-Appliance	Hardware-Appliance	Hardware-Appliance	Hardware-Appliance
Basisbetriebssystem	Linux	SonicOS	Windows	Linux from Scratch	Linux	Fireware	ZLD (linuxbasiert)
Features							
zusätzliche Services zu FW, IDS/IPS, Antivirus	VPN, Content-Filter, Antispam u. a.	VPN, Content-Filter, Antispam	Endpointprotection (Anwendungs-/ Prozesskontrolle) sofern zusätzlicher Client installiert wird	VPN, Content-Filter Antispam u. a.	VPN, URL-Filter, Content-Filter, Antispam u. a.	URL-Filter, Antispam	VPN, Content-Filter Application Patrol, Antispam
Integration bestehender Produkte	von Mailshell	–	eigene, Sygate	Kaspersky	eigene, Trend Micro	eigene, Surfcontrol, Commtouch	eigene
QoS (Zusich. v. Bandbreite)	✓	✓	–	✓	✓	✓	✓
Failover	✓	nur PRO 3060	✓	aktiv/passiv	✓	je nach Produkt aktiv/aktiv, aktiv/passiv	✓
Clustering	✓	ab PRO 3060	✓	✓	–	–	–
VLAN-Unterstützung	✓	✓	–	✓	✓	–; ✓; ✓	✓
Erweiterungsmöglichkeiten	✓	–	✓	Interfaces, Kryptobeschleuniger, MINI GBIC	✓	durch Software Keys	✓
Layer 7 Inspection/ für welche Protokolle und Dienste	–	Application Firewall	✓	SMTP, POP3, HTTP, FTP, P2P	SMTP, HTTP, FTP, SIP	SMTP, POP3, HTTP, FTP, DNS	✓ (für IDS/IPS)
unterstützte Routing-Protokolle	RIP	RIP, OSPF	–	RIP, RIPv2, OSPF, BGP	–	OSPF, RIP, RIPv2, BGP	RIP, OSPF
Performance							
Anzahl d. Interfaces	6	8; 6	k. A.	3 – 6	bis 16	6; 8; 8	7
Beschränk. IP-Adressen	–	–	✓	–	durch Lizenz	–	–
max. Anzahl Verbindungen pro Sek., alle Dienste aktiv	500	25 000; k. A.	abhängig vom Betriebssystem	> 1 500 000	limitiert durch Parameter	k. A.	60 000
max. Anz. paralleler Verbindungen, alle Dienste aktiv	1000	je nach Gerät	abhängig vom Betriebsystem	1 500 000	100 000	10 000; 200 000; 1 000 000	60 000
max. Durchsatz, alle Dienste aktiv	933 MBit/s	1,2 GBit/s; 25 MBit/s	abhängig vom Betriebssystem	2,4 GBit/s (FW), 540 MBit/s (FW+IDS), 600 MBit/s (VPN)	100 MBit/s	100 MBit/s; 1,5 GBit/s; 2,3 GBit/s	200 MBit/s (FW), 48 MBit/s (FW+UTM), 100 MBit/s (FW+VPN)
Management							
Management über zentr. Interface	✓	✓	✓	–	✓	✓	✓
Authentifizierungsmöglichkeiten	LDAP, RADIUS, AD u. a.	LDAP, RADIUS, AD, lokal	LDAP, AD, Secure ID	LDAP, RADIUS, AD, lokal	LDAP, RADIUS, AD	LDAP, RADIUS, AD, Secure ID	LDAP, RADIUS, AD, lokal
Benutzer- u. Rollenkonzept	✓	✓	✓	–	✓	–; ✓; ✓	✓
Dienste unabhängig auf versch. Interfaces einsetzbar	✓	k. A.	✓	✓	✓	–; ✓; ✓	✓
Reihenfolge der Dienste definierbar	–	–	–	–	–	–	–
Logging/Alarming/Reporting							
wie erfolgt Logging/ erlaubte Dateigröße	k. A./max. Festplattengröße	Syslog oder proprietär	Syslog	Syslog	in die Datenbank	Syslog oder proprietär	Syslog/ begrenzte Anzahl Messages
welche Alarmmöglichkeiten	SNMP, SMS, E-Mail	SNMP, E-Mail	SMTP, Pop-up-Fenster	E-Mail	E-Mail, Web-Tool	SNMP, E-Mail, Pop-up-Fenster	SNMP, E-Mail
Log-Korrelation	✓	k. A.	✓	–	–	✓	–
Generieren von Reports möglich	✓	Global Mgmt. System, ViewPoint	SQL-Abfragen über Wizard	✓	✓	Reporting-Tool	✓
Sonstiges							
Preis (netto)	4000 €	24 995 $; 2795 $	44 € pro Lizenz bei 100 Liz. (Bsp.)	ab 695 €	4000 € für 50 Benutzer (Bsp.)	ab 528 $; ab 1605 $; ab 6590 $	1250 €
Anmerkungen	–	Multi-Core-Architektur; k. A.	–	–	–	–	–

Tabelle beruht auf Herstellerangaben
✓ ja/zutreffend/vorhanden
– nein/nicht zutreffend/nicht vorhanden
k. A. keine Angabe

allerdings nicht blind auf die Angaben der Hersteller verlassen. Denn diese sind kaum miteinander vergleichbar, da jeder Hersteller seine eigene Methodik für die Messungen besitzt. Bei einer Vorauswahl sollte man vor allem darauf achten, dass die Performance-Messungen bei Aktivierung aller Dienste einer Appliance erfolgen. Je nachdem, wie viele Dienste in einem einzelnen System integriert sind und wie stark der Messdatenstrom die einzelnen Dienste in Anspruch nimmt, sind Performance-Einbußen von 60 – 80 % im Durchsatz keine Seltenheit.

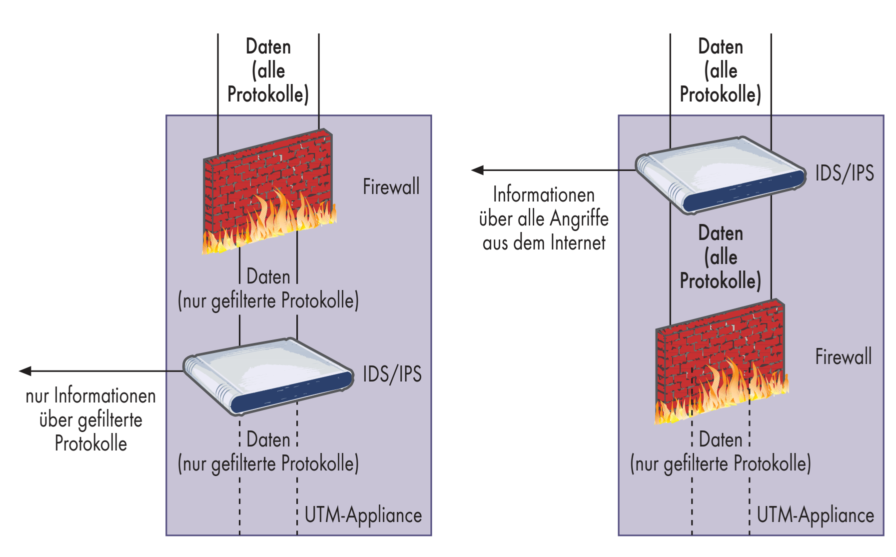
Wenn möglich, sollte man in Bezug auf den Durchsatz eigene Tests durchführen. Dabei sollte der Messdatenstrom vergleichbar mit dem Produktivdatenstrom und alle für den Produktivbetrieb notwendigen oder geplanten Dienste aktiviert sein. Ein Datenstrom, der etwa zur Hälfte aus großen Dateien besteht, die via VPN übertragen und anschließend auf Viren geprüft werden müssen, ergibt andere Performance-Werte als einer, der zu 80 % aus Internetzugriffen besteht und lediglich aktivierte URL-, Content Filter und Virenscanner benötigt.

Ebensowenig sollte man vergessen, dass jedes Gerät lediglich eine maximale Anzahl paralleler Verbindungen verarbeiten kann. Im Allgemeinen sind das bei den meisten teuren Geräten weit mehr als 100 000, aber auch hier ist eine Prüfung mit produktivnahen Daten sinnvoll. Insbesondere dann, wenn Verbindungen dauerhaft oder über längere Zeit gehalten werden müssen (zum Beispiel VPN, Online-Interfaces zu bestimmten Anwendungen et cetera).

Hohe Performance benötigt

Will man Dienste nach außen anbieten, sollte man darauf achten, wie schnell ein Gerät beim Verbindungsaufbau ist, sprich wie viele Verbindungen es pro Sekunde aufbauen kann. Ein stark frequentierter Webserver hat hier sicherlich andere Anforderungen an ein UTM-Produkt als ein System, das etwa als VPN-Access-Point für die Sales-Mannschaft eines Unternehmens zuständig ist.

Doch die Performance eines Systems ist nur ein Teilaspekt. Insbesondere große Unternehmen sollten auf die Management-Möglichkeiten von Produkten Wert legen. Hatten die meisten Systeme aus der Anfangszeit lediglich ein Interface, das mehr oder weniger schick die einzelnen Funktionen miteinander verband und direkt angesprochen werden musste, so existiert heute für nahezu jedes Produkt eine gemeinsame Management-Oberfläche, die es ermöglicht, Policies zentral abzulegen und auf mehrere Systeme zu verteilen.



Je nach Reihenfolge der Dienste erzielt man unterschiedliche Ergebnisse: Bei vorgeschalteter Firewall (Variante links) verarbeitet das UTM-Gerät die gefilterten Daten performanter, bei vorgeschaltetem IDS/IPS (rechts) erhält man mehr Informationen über Angriffe aus dem Internet (Abb. 1).

Je nach Gerät und Hersteller muss man eine solche Management-Software zusätzlich kaufen – doch daran sollte im Interesse des effektiveren Arbeitens kein IT-Leiter sparen. Das Erstellen von eigenen Skripten zur Verteilung von Policies oder der Aufbau eigener Datenbanken zur Versionskontrolle et cetera sind zeitaufwendige Vorgänge, die die Kosten für ein zu kaufendes Tool meist schnell übersteigen. Bei einigen auf dem Markt befindlichen Geräten (s. Tabelle) werden sich Administratoren jedoch mit eigenen Werkzeugen behelfen müssen.

Durch die Verbindung mehrerer Dienste in einem Gerät kann es zu Komplikationen bei der Administration dieser Systeme kommen. So ist es in Unternehmen nicht selten, dass unterschiedliche Gruppen für Virenschutz und Firewalls zuständig sind. Ist das der Fall, so sollte man unbedingt auf ein entsprechendes Benutzer- und Rollenkonzept achten. Im Mittelpunkt steht die Frage, wer mit welchen Rechten auf welchen Bereich zugreifen darf. Bei nahezu allen aufgeführten Produkten ist es möglich, die Benutzeroberflächen so zu konfigurieren, dass einzelne Bereiche entweder vollständig ausgeblendet sind oder zumindest der Zugriff auf sie nur lesend gestattet ist.

Neben der Einrichtung des Zugriffschutzes auf dem UTM-Gerät ist es sinnvoll, ihn in das zentrale Benutzermanagement eines Unternehmens einzubinden. Alle Hersteller bieten diese Option zumindest über LDAP. Wer seine UTM-Appliance in das Active Directory seines Unternehmens direkt einbinden möchte, wird bei immerhin 15 Herstellern fündig. Stärkere Authentifizierungsmechanismen wie SecureID bieten fünf Hersteller an. Die Integration über Protokolle wie RADIUS ist bei zahlreichen Produkten möglich. Sinnvoll ist die Integration in die zentrale Benutzerdatenbank immer dann, wenn es einen definierten Prozess für den Eintritt und das Ausscheiden von Mitarbeitern gibt.

Bei der Auswahl eines UTM-Produktes nicht unerheblich ist die Frage, wie fein abgestuft man es konfigurieren kann. Denn unter Umständen kann es die Performance maßgeblich beeinflussen, wenn die angebotenen Dienste unabhängig auf verschiedenen Interfaces einsetzbar sind. Ob etwa ein Virenscan (E-Mail und HTTP) auf allen Interfaces durchgeführt werden muss oder nur den Datenstrom von außen ohne intern verschickte E-Mails berücksichtigt, macht einen großen Unterschied.

Analog dazu stellt sich die Frage, auf welchen Interfaces ein IDS/IPS lauschen soll. Reichen hier einzelne Bereiche aus oder sollen sie alles prüfen – eine deutlich andere Rechenleistung. Glücklicherweise bieten fast alle Hersteller die Möglichkeit, eine solche dedizierte Konfiguration vorzunehmen. Anders sieht es aus, wenn man die Reihenfolge der Dienste beeinflussen will. Es ist beispielsweise ein großer Unterschied, ob der Datenstrom an einem externen Interface zuerst von der eingebauten Firewall oder vom IDS/IPS geprüft wird (Abb. 1). Im ersten Fall betrachtet das IDS/IPS lediglich einen Bruchteil der Daten. Manchmal kann es jedoch sinnvoll sein, den Datenstrom vor einer Firewall zu überwachen. Das liefert Daten (und

teilweise auch Argumentationshilfen) über die reale Anzahl von Angriffsversuchen aus dem angebundenen Netz. Lediglich Secure Computing gibt an, diese Wahlmöglichkeit im Produkt verankert zu haben. Alle anderen Hersteller bearbeiten den Datenstrom in einer vorgegebenen Reihenfolge.

Aufbau und Konfiguration von Sicherheitssystemen sind nur zwei der Aufgaben eines Sicherheits-Administrators. Das Logging, Alarming und Reporting sind meist vernachlässigte, aber dennoch sehr wichtige Komponenten, die das eigene Netzwerk absichern. Bei UTM-Systemen kommt zudem die Besonderheit hinzu, dass es nicht nur einen, sondern gleich mehrere unterschiedliche Dienste zu betrachten gilt. Logfiles sollten die UTM-Systeme daher in einem Format schreiben, das sowohl leserlich ist als auch durch weiterführende Anwendungen oder Systeme bearbeitet (Alarming, Reporting) werden kann.

Gut geloggt ist halb gewonnen

Die meisten Hersteller lösen dieses Problem, indem sie das Logging via Syslog durchführen. Teilweise existieren zusätzliche proprietäre Formate, die die hauseigenen zentralen Management-Tools verarbeiten können. Lediglich Vasco schreibt die Logdaten in eine eigene Datenbank und liefert sie nicht via Syslog weiter. Check Point schreibt die Logdaten ebenfalls zuerst in einem proprietären Format auf den zugehörigen Managementserver, von dort ist allerdings eine Weiterleitung via Syslog möglich.

Für das Weiterverarbeiten von Logdaten bieten die meisten Hersteller eigene Reporting-Tools an, die teilweise zusätzlich zu erwerben sind. Für die im Syslog-Format vorliegenden Logdaten kann man auch eigene Werkzeuge verwenden. Allerdings ist hier Vorsicht geboten, da es einen hohen Konfigurationsaufwand nach sich zieht, wenn man gute Ergebnissen erhalten möchte. Anzumerken ist außerdem, dass bei den Geräten nicht alle Syslognachrichten, die versendet werden, dokumentiert sind. Eine Auswertung der Logfiles kann daher mühevoll sein, wenn man nicht weiß, was man zusammenfassen kann oder nach welchen Nachrichten man suchen muss.

Die Integration mehrer Dienste in einer Appliance könnte nahelegen, dass die meisten Hersteller das genutzt haben und für das Alarming eine Korrelation der Logfiles durchführen. Das würde bedeuten, dass ein UTM-Produkt die Logdaten eines Dienstes nicht isoliert betrachtet, sondern das System in seiner Gesamtheit (Abb. 2). Ein Portscan von einer bestimmten IP-Adresse aus ist nicht zwingend ein Grund, Alarmstufe Rot auszulösen. Wenn aber die gleiche IP-Adresse zusätzliche Warnmeldungen im IDS/IPS generiert und der Virenscanner ebenfalls bei dieser IP anschlägt, könnte eine Benachrichtigung des Administrators durchaus angebracht sein.

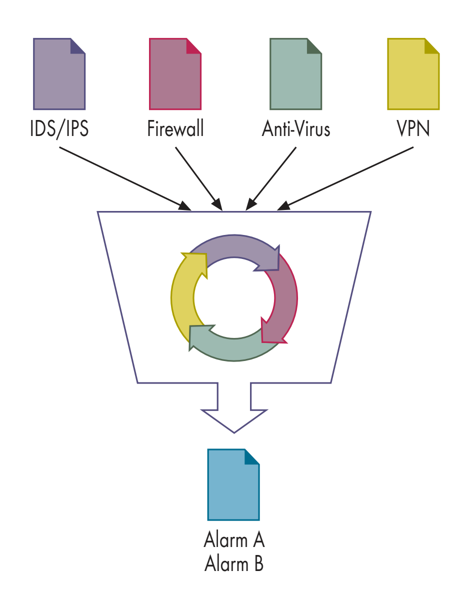
Das Auslösen eines Alarms bei einzelnen Ereignissen bietet jeder Hersteller in der einen oder anderen Form (SNMP, E-Mail, Syslog et cetera). Bei einigen Produkten kann man die Alarme relativ frei auf der Basis von Logdaten definieren, andere benötigen hierzu wiederum die zusätzlich zu erwerbende Managementsoftware. Ein wirkliches Log-Event-Correlation-Tool wie das Open-Source-Werkzeug „Simple Event Correlator“ oder kommerzielle sogenannte Security-Information-Management-Werkzeuge (SIM), bietet kaum ein Hersteller – es sei denn mit zusätzlichen Produkten. Das wäre jedoch ein ausgesprochen sinnvolles Feature, zumal schließlich alle Anwendungen bereits unter einem Dach vereint sind und es somit nur ein weiterer logischer Schritt bei der Weiterentwicklung von UTM wäre.

Individuelle Tests

Seit die ersten UTM-Systeme auf den Markt gekommen sind, hat sich einiges getan. Nahezu jedes Produkt wurde um zahlreiche Funktionen erweitert, und ein Ende dieser Entwicklung ist noch nicht abzusehen. Ob durch die Fülle der Funktionen und deren Integration in eine Benutzerumgebung ein einfaches Arbeiten möglich bleibt – einer der bislang bei UTM propagierten Vorteile – wird sich zeigen. Der Einsatz von Einzelprodukten ist jedoch nicht zwingend die einfachere Alternative.

Schließlich muss jeder Administrator oder IT-Leiter selbst entscheiden, welche Features er als relevant oder obsolet erachtet. Wichtig ist es vor allem, das Ziel, das man mit einem solchen Gerät erreichen will, nicht aus den Augen zu verlieren. Ein kleines oder mittelständisches Unternehmen benötigt mit Sicherheit kein Benutzer-Interface, mit dem man mehrere Hundert Geräte managen kann.

Auch ist fraglich, ob es nicht in wirklich großen Umgebungen doch sinnvoller ist, dedizierte Geräte für rechenintensive Aufgaben wie VPN et cetera einzusetzen. Allerdings besitzen die aktuellen UTM-Systeme genügend Features, die sie insbesondere für große Unternehmen interessant machen. Die Anbindung von kleinen sowie größeren Außenstellen durch ein zentrales Management und stetige Erweiterung dieser Produkte ist dabei nur ein Aspekt.



Idealerweise sollten UTM-Systeme ebenso wie Security-Information-Management-Systeme aus allen Logs die Daten korrelieren und bei drohender Gefahr Alarm auslösen – dieses Feature ist bei heutigen Produkten aber in der Regel noch eine Wunschvorstellung (Abb. 2).

Wichtig ist jedoch, dass vor dem Kauf eines solchen Produktes ein ausgiebiger Test steht. Durch die vielen Dienste und die Einsatzszenarien (und damit auch Datenströme), die in jedem Unternehmen anders sind, können die einzelnen Produkte ganz unterschiedlich reagieren. Die Performance-Angaben der Hersteller geben zwar eine grobe Orientierung, machen aber einen Test mit sinnvollen Daten, wie sie im Unternehmensalltag vorkommen, nicht obsolet.

Darüber hinaus ist zu beachten, dass zwar nur ein Gerät im Test steht, dieses aber eine große Anzahl von Features enthält. Deswegen sollte man die Testdauer nicht zu kurz ansetzen. Der Test einer Firewall, eines IDS/IPS, einer Anti-Viren-Software, eines VPNs, eines URL- oder Content Filters, einer Anti-Spam-Software et cetera kann nun mal nicht in einer Woche abgehakt werden. Das sollte jedem IT-Leiter klar sein. (ur)

JÖRN MAIER

ist seit über 6 Jahren im Bereich IT-Sicherheit als Berater tätig.





Anomalien im Netz via RRD-Tool und Cricket erkennen

Immer einen Schritt voraus

Pascal Schöttle, Thomas Götz

Für die manuelle Überwachung der Netzauslastung muss ein Unternehmen erhebliche personelle Ressourcen seiner IT-Abteilung binden. Eine Kombination des RRD-Tools mit dem Holt-Winters-Algorithmus verspricht Entlastung – und eine frühzeitige Erkennung von Netzanomalien.

Bei der Vielzahl von Netzwerk-Interfaces, schon in kleineren Unternehmen, lässt sich die Auslastung nicht mehr bei allen Anschlüssen manuell überwachen. Von großem Vorteil wäre ein Tool, das dem Netzadministrator automatisch eine Vorauswahl präsentiert, die er detaillierter überprüfen kann. Einfache Schwellwerte reichen dabei als Auswahlkriterium jedoch oft nicht aus, da sie nur einen kleinen Teil der potenziellen Probleme abdecken. Dank dynamisch angepasster Schwellwerte reagiert der in diesem Artikel vorgestellte Holt-Winters-Algorithmus (siehe Kasten) auf Anomalien, die nicht nur aus der Überlastung der Netzanbindung herrühren. Als Beispiele lassen sich außerplanmäßige größere Datenübertragungen oder eine reduzierte Internet-Bandbreite aufgrund von Schwierigkeiten des Providers anführen. Auch Angriffe von innerhalb oder außerhalb des Netzes stellen eine Anomalie dar, die der Holt-Winters-Algorithmus erkennen kann.

Sammelstelle für Messdaten

Als Datensammler kommt in diesem Szenario das von Tobi Oetiker entworfene und für grafische Auswertungen verbreitete RRD-Tool (siehe „Onlinequellen") zum Einsatz. RRD steht dabei für die sogenannten Round-Robin-Datenbanken, in denen sich unterschiedlichste Messdaten speichern lassen. Die RRDs erzeugt man mit einer festen Größe und einer festen Anzahl von Werten. Sobald mehr Messwerte vorliegen, werden einfach die ältesten Werte überschrieben. So besteht keine Gefahr, dass die RRDs zu stark wachsen.

Jede RRD ist in weitere Round-Robin-Archive (RRA) unterteilt, in denen man die Messdaten – auch in unterschiedlicher Granularität – speichern kann. So ist es beispielsweise sinnvoll, die Messdaten im 5-Minuten-Takt für 2 Tage zu speichern, die im 30-Minuten-Takt jedoch für 2 Wochen oder im 2-Tages-Takt für ein ganzes Jahr. Damit hat der Administrator sowohl einen detaillierten Überblick, was in letzter Zeit im Netz passiert ist, als auch gleichzeitig eine Übersicht über einen längeren Zeitraum. Neben der effizienten Speicherung der Daten bietet RRD-Tool auch Mechanismen zur grafischen Darstellung der gesammelten Werte.

Um nun den Holt-Winters-Algorithmus im RRD-Tool nutzen zu können, gibt der Nutzer bei der Initialisierung der RRD-Dateien, neben den RRAs für die verschiedenen Taktungen, fünf zusätzliche RRAs für den Vorhersagealgorithmus an. Das erste speichert die vorhergesagten Werte, das zweite die saisonale Komponente, das dritte die saisonale Abweichung und das vierte die vorhergesagte Abweichung. Im fünften RRA merkt sich das Tool, ob eine Anomalie vorliegt oder nicht. Die RRAs für die Anomalieerkennung legt man genau wie normale RRAs mit dem Befehl *rddtool create* an, beispielsweise:

```
rrdtool create anomalie.rrd -- step 300 ↙
  DS:ifHCInOctets:COUNTER:600:0:U ↙
  DS:ifHCOutOctets:COUNTER: ↙
                            600:0:U ↙
  RRA:AVERAGE:0.5:1:4032 ↙
  RRA:HWPREDICT:2016:0.1: ↙
                  0.0035:288:3 ↙
  RRA:SEASONAL:288:0.1:2 ↙
  RRA:DEVSEASONAL:288:0.1:2 ↙
  RRA:DEVPREDICT:2016:4 ↙
  RRA:FAILURES:288:9:15:4
```

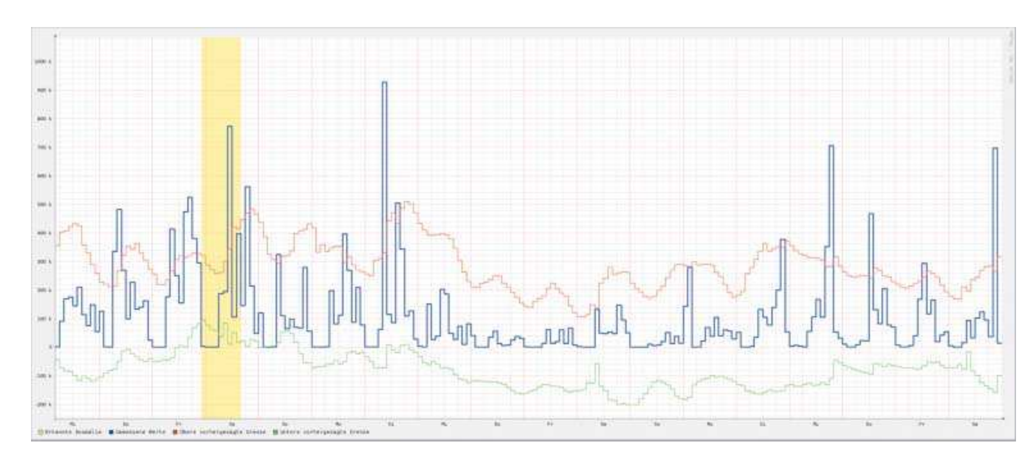
Dieser Aufruf erzeugt eine RRD-Datei namens *anomalie.rrd*, die alle 5 Minuten (*--step = 300 s*) einen Wert speichert. In ihr sind zwei Datenquellen definiert. Die erste speichert die eingehenden und die zweite die ausgehenden Bytes; beide sind vom Typ *COUNTER*. Falls 10 Minuten (= 600 s) kein Wert gemessen werden kann, setzt RRD-Tool den Messwert auf *Unknown*. Als Nächstes wird das RRA für den Durchschnittswert (*AVERAGE*) der letzten 5 Minuten definiert. Hier werden Daten für zwei Wochen (= 4032 Beobachtungen à 5 Minuten) gespeichert. Darauf folgt das RRA mit den Vorhersagen (*HWPREDICT*). Das angegebene Beispiel speichert Vorhersagen für einen Zeitraum von einer Woche ((7×24×12) = 2016 Beobachtungen à 5 Minuten), setzt den Parameter α auf 0.1, den Parameter β auf 0.0035, die Periode auf einen Tag (288 Beobachtungen à 5 Minuten) und gibt als Letztes noch die Nummer des *SEASONAL-RRA* an. Die beiden RRAs *SEASONAL* und *DEVSEASONAL* haben als Parameter die Periode (hier 1 Tag), den saisonalen Glättungsparameter γ (hier 0.1) und die Nummer des *HWPREDICT-RRA*.

Das RRA DEVPREDICT verfügen als ersten Parameter über die Anzahl der zu speichernden Werte und als zweiten die Nummer des DEVSEASONAL-RRA. FAILURES-RRA besitzt als ersten Parameter wieder die Periode, hier gleichbedeutend mit der Anzahl der zu speichernden Werte, als zweiten die Schranke s, ab der Alarm geschlagen werden soll (hier 9), als dritten die Länge des zu berücksichtigenden Zeitfensters l (hier 15) und als letzten Parameter wieder die Nummer des DEVSEASONAL-RRA. Hierbei sollte man beachten, dass der höchste Wert, den RRD-Tool für die Länge des Zeitfensters verwenden kann, 28 beträgt (per Definition im RRD-Tool). Im FAILURES-RRA steht entweder eine „0“, wenn der Algorithmus keine Anomalie erkannt hat, oder eine „1“, wenn in den letzten l Beobachtungen mindestens s Verletzungen der vorhergesagten Schranken aufgetreten sind. Die Schranken berechnen sich wie im Kasten „Holt-Winters-Algorithmus“ beschrieben.

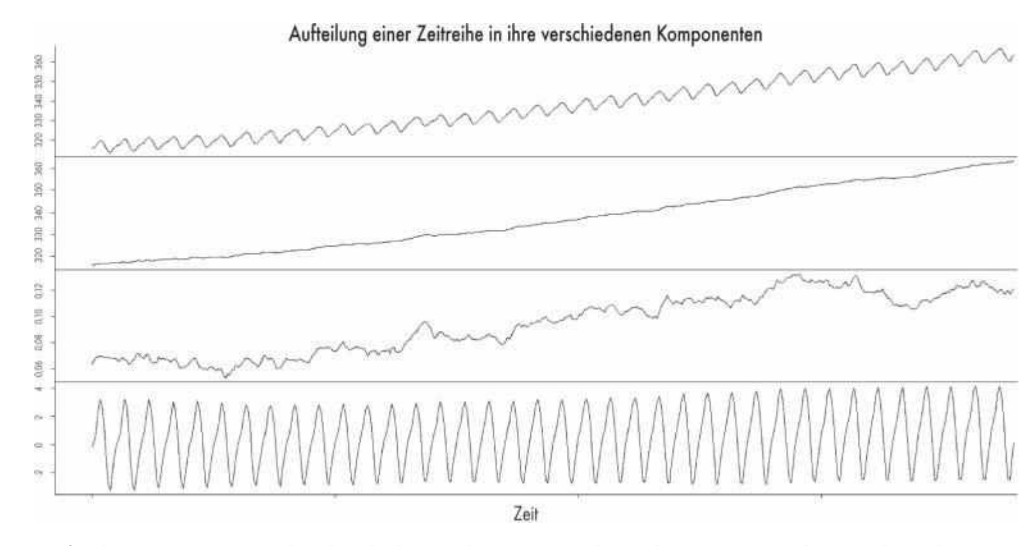
Die beiden Parameter δ_- und δ_+ lassen sich nachträglich mit dem Befehl

```
rrdtool tune anomalie.rrd --deltapos ↙
    VALUE+ --deltaneg VALUE
```

auf den Wert *VALUE+* beziehungsweise *VALUE-* verändern. Ihr Default-Wert beträgt jeweils 2.0. Auch wenn man den Wert theoretisch beliebig ändern kann, hat sich in der Praxis herausgestellt, das ein Wert im Intervall [2, 3] sinnvoll ist. Um die RRD-Datei nun auf dem Laufenden zu halten, verwendet man am besten eines der vielen frei



***rrdtool*-Ausgabe: Die grüne Linie zeigt die untere und die rote Linie die obere vorhergesagte Grenze; die tatsächlich gemessenen Werte sind blau. Der gelb eingefärbte Hintergrund hebt eine Anomalie hervor. In diesem Beispiel beträgt die Länge des Zeitfensters 9 Beobachtungen und als Anomalie gelten 5 Verletzungen der Grenzen in dieser Zeit (Abb. 1).**



Aufteilung einer Zeitreihe durch den Holt-Winters-Algorithmus: Ganz oben stehen die vorhergesagten Werte, gefolgt von der Baseline. Im dritten Abschnitt sieht man den langfristigen Trend und ganz unten die saisonale Komponente. Die vorhergesagten Werte bilden sich aus der Summe der drei anderen Komponenten (Abb. 2).

iX-TRACT

- Netzadministratoren müssen viel Zeit in die Überwachung von Netzanschlüssen investieren – selbst wenn sie sich auf die zentralen Server beschränken.
- Auf Basis gemessener Auslastungswerte lassen sich mit dem Holt-Winters-Algorithmus Vorhersagen für zukünftige Messpunkte treffen.
- Mit Holt-Winters in Kombination mit dem RRD-Tool und dessen GUI-Frontend Cricket kann sich der Netzverwalter eine Menge Arbeit ersparen.

Listing 1:monitor-thresholds definieren

```
target --default--
mail-pgm = /usr/bin/mailx
persistent-alarm=false
monitor-thresholds =
"ifHCInOctets : failures : 0 :n: FILE : /var/logs/cricket-alerts,
ifHCInOctets : failures : 0 :n: MAIL : %mail-pgm% : me\\\@mydomain.de,
ifHCOutOctets : failures : 0 :n: FILE : /var/logs/cricket-alerts,
ifHCOutOctets : failures : 0 :n: MAIL : %mail-pgm% : me\\\@mydomain.de"
```

erhältlichen Frontends wie Cricket.

Alles vereint in Cricket

Cricket ist ein Frontend zum RRD-Tool, mit dem sich die verschiedenen RRDs verwalten und daraus automatisch eine HTML-Grafik der gemessenen Werte erzeugen lassen. Cricket entstand unter der Regie von Microsofts Web-TV – mittlerweile MSN – und ist frei erhältlich (siehe „Onlinequellen“). Im Wesentlichen besteht es aus zwei Teilen. Dem Datensammler, der alle fünf Minuten (per Default) die vorher erzeugten RRD-Dateien aktualisiert, und einer via Web-Interface ansprechbaren Grafikausgabe. In den Konfigurationsdateien lässt sich angeben, mit welchen Parametern die RRD-Dateien erzeugt werden sollen. Das oben angegebene Beispiel würde in einer Cricket-Konfiguration folgendermaßen aussehen:

```
rra 5minAve AVERAGE:0.5:1:4032
rra hwpredict HWPREDICT:2016:0.1:0.0035:288:3
rra seasonal SEASONAL:288:0.1:2
rra devseasonal DEVSEASONAL:288:0.1:2
rra devpredict DEVPREDICT:2016:4
rra failures FAILURES:288:9:15:4
```

Nun kann man sich die gemessenen und vorhergesagten Werte über die grafische Ausgabe anzeigen lassen. Dabei kann man zwischen drei Optionen wählen: nur die gemessenen Werte, die gemessenen Werte zusammen mit den vorhergesagten Grenzen oder die Werte mit den Grenzen und den Anomalien, sofern aufgetreten. Damit der Administrator nicht ständig auf den Monitor achten muss und trotzdem keine Anomalie verpasst, verfügt Cricket über einige Alarmierungsmöglichkeiten, die greifen, sobald das *FAILURES-RRA* in der RRD-Datei eine „1“, also eine erkannte Anomalie, enthält. Als Alarmierung kann man eine der folgenden Optionen wählen: *SNMP* generiert einen SNMP-Trap. *MAIL* verschickt bei einem Alarm und bei einer Entwarnung eine

Holt-Winters-Algorithmus

Der Vorhersagealgorithmus nach Holt & Winters basiert auf der mathematischen Methode des exponentiellen Glättens, einem Verfahren zur Analyse und kurzfristigen Prognose von Zeitreihen. Exponentielles Glätten ist ein einfacher Algorithmus zur Vorhersage des nächsten Werts $\hat{y}(t+1)$ einer Zeitserie y, wenn der aktuell gemessene Wert $y(t)$ und der aktuell vorhergesagte Wert $\hat{y}(t)$ vorliegen. Es gilt dann:

$$\hat{y}(t+1) = \alpha y(t) + (1-\alpha)\hat{y}(t)$$

Der Name exponentielles Glätten kommt zustande, da weiter zurückliegende Werte immer weniger ins Gewicht fallen. Die Gewichtung ist exponentiell fallend. So ist beispielsweise:

$$\hat{y}(t) = \alpha y(t-1) + (1-\alpha)\hat{y}(t-1)$$

und somit: $\hat{y}(t+1) = \alpha y(t) + (1-\alpha)\alpha y(t-1) + (1-\alpha)\hat{y}(t-1)$.

Wenn man die $\hat{y}$ weiter ersetzt, kommt man zu folgendem Ergebnis:

$$\hat{y}(t+1) = \alpha y(t) + \alpha(1-\alpha)y(t-1) + \alpha(1-\alpha)^2 y(t-2) + \ldots + \alpha(1-\alpha)^{t-1}y(1) + \alpha(1-\alpha)^t y(0)$$

Hierbei ist $0 < \alpha \leq 1$ der sogenannte Glättungsparameter, der angibt, wie viel Gewicht der zuletzt gemessene Wert bekommt. Bei $\alpha = 1$ würde nur der letzte Wert zählen, und je näher α an 0 ist, umso mehr Gewicht fällt auf die älteren Beobachtungen.

Im Kern nutzt der Holt-Winters-Algorithmus zwar das exponentielle Glätten, berücksichtigt jedoch sich periodisch wiederholende Ereignisse (mit Periode p) und einen langfristigen Trend in der beobachteten Zeitreihe. Dazu spaltet er die Zeitreihe $y(t)$ in drei Komponenten: die Baseline $a(t)$, den linearen (langfristigen) Trend $b(t)$ sowie die saisonale Komponente $c(t)$.

Die eigentliche Zeitreihe $y(t)$ setzt sich aus diesen, jeweils exponentiell geglätteten, Komponenten zusammen, also $y(t) = a(t-1) + b(t-1) + c(t-p)$.

Zur Berechnung von $y(t)$ benötigt man $a(t-1)$ und $b(t-1)$ deshalb, weil die Berechnung von $a(t)$ den aktuell gemessenen Wert $y(t)$ benutzt und die Berechnung von $b(t)$ den Wert von $a(t)$ (siehe Formeln weiter unten).

Der Holt-Winters-Algorithmus hat zum Ziel, zukünftige Werte $\dot{y}(t)$ der Zeitreihe y vorherzusagen. Dafür benötigt man eine Zeitreihe von mindestens $2 \times p$ Länge (Mindestlänge für das Baselining). $\hat{y}(t+1)$ ist dann die Summe aus den einzelnen Komponenten. Natürlich werden die einzelnen Teile in jedem Schritt neu angepasst. Da der Holt-Winters-Algorithmus es nicht berücksichtigt, kann man manuell noch triviale Anpassungen vornehmen. So kann beispielsweise im Einsatzszenario des Artikels der Netzverkehr nie unter Null fallen. Da der Algorithmus hauptsächlich saisonale Steigungen respektive Gefälle berücksichtigt, kann es passieren, dass in einer Vorhersage der Netzverkehr unter Null fallen würde, was wenig sinnvoll ist.

Formeln für den Holt-Winters und seine Parameter:

Baseline

$$a(t) = \alpha(y(t)-c(t-p)) + (1-\alpha)(a(t-1) + b(t-1))$$

Linearer Trend

$$b(t) = \beta(a(t)-a(t-1)) + (1-\beta)(b(t-1))$$

Saisonale Komponente

$$c(t) = \gamma(y(t)-a(t)) + (1-\gamma)(c(t-p))$$

Also: $c(t-p) = \gamma(y(t-p)-a(t-p)) + (1-\gamma)(c(t-2p))$ und somit

$$\hat{y}(t+1) = a(t) + b(t) + c(t+1-p)$$

α, β und γ sind hierbei die Glättungsparameter. Sie liegen immer im Intervall [0,1] und geben an, wie viel Gewicht der letzte gemessene Wert bekommen soll. Sind sie groß (~1) reagiert der Algorithmus stark auf den letzten Messwert. Daraus folgt eine schnelle Anpassung an plötzliche Veränderungen, aber kaum Berücksichtigung der älteren Messwerte. Sind sie jedoch klein (~0), reagiert er nur sehr schwach auf den letzten Messwert. Hieraus folgt eine langsamere Anpassung, dafür aber eine höhere Gewichtung der älteren Messwerte.

Anomalie

Zum Erkennen von Anomalien bildet man ein Konfidenzintervall mit einer unteren und einer oberen Schranke. Dazu definiert man die sogenannte Messabweichung $d(t)$, welche der gewichtete Durchschnitt der absoluten Abweichung ist. In sie fließen sowohl gemessene als auch vorhergesagte Werte mit ein.

$$d(t) = \gamma|y(t)-\hat{y}(t)|+(1-\gamma)d(t-p)$$

Außerdem benötigt man zwei Parameter δ_+ und δ_-. Aus d(t-p), δ_+ und δ_- bildet man dann das Konfidenzintervall für $y(t)$:

$$(\hat{y}(t)-\delta_- d(t-p), \hat{y}(t) + \delta_+ d(t-p))$$

Als Anomalie gelten jetzt nicht einzelne Verletzungen dieses Intervalls, da dies eine zu hohe Anzahl von False Positives hervorrufen würde. Deshalb arbeitet die Anomalie-Erkennung mit einem festen Zeitfenster der Länge l und einer oberen Schranke s. Wenn die Anzahl der Verletzungen des Intervalls innerhalb der letzten l Beobachtungen s erreicht, gilt dies als eine Anomalie. Beispielsweise registriert der Algorithmus in der Konfiguration des Haupttextes eine Anomalie, falls in den letzten $l = 15$ Beobachtungen $s = 9$ Verletzungen des Intervalls aufgetreten sind.

E-Mail an die angegebene Adresse. *FILE* schreibt bei einem Alarm eine Zeile in die angegebene Datei und löscht diese wieder, falls der Alarm vorbei ist. *EXEC* führt das angegebene Shell-Kommando oder -Skript aus. *FUNC* arbeitet ähnlich wie *EXEC*, nur dass es ein Perl-Skript ausführt. *META* schließlich speichert den Alarm in den Cricket-eigenen Meta-Dateien.

Um diese Alarmierungsmöglichkeiten einsetzen zu können, muss man in der Konfigurationsdatei entsprechende *monitor-thresholds* definieren. Listing 1 zeigt, wie das für den hier beschriebenen Fall aussehen könnte. Zur Erläuterung: *mail-pgm* gibt den Pfad zum verwendeten E-Mail-Programm an. Die Option *persistent-alarm* legt fest, ob die definierte Aktion bei jedem neuen Alarm erfolgt (*true*), oder nur, falls es zwischen zwei Alarmen wieder eine Entwarnung gegeben hat (*false*). Der erste Wert bei *monitor-thresholds* gibt die Datenquelle an. Das folgende *failures* stellt ein, dass *FAILURES-RRA* als Informationsquelle heranzuziehen ist, gefolgt vom Bereich (des Messwerts), in dem Alarm geschlagen werden soll (*n* bedeutet keine obere beziehungsweise untere Grenze). Im letzten Part findet sich die Alarmierungsart mit den dazugehörigen Parametern. Im Beispiel schreibt Cricket sowohl beim ein- als auch beim ausgehenden Verkehr den Alarm in die Datei */var/logs/cricket-alerts*, respektive löscht ihn, wenn der Alarm vorbei ist, wieder daraus. Darüber hinaus erhält *me@mydomain.de* bei jedem Alarm und jeder Entwarnung eine E-Mail.

Fazit

Mit RRD-Tool, Cricket und dem Holt-Winter-Algorithmus hat man eine relativ einfach zu konfigurierende und somit relativ schnell einsatzbereite Open-Source-Anwendung, um Anomalien zu erkennen. Der Holt-Winter-Algorithmus liefert vor allem in den Bereichen ein respektables Ergebnis, in denen man periodisch wiederkehrende Muster hat. Der Netzwerkverkehr über einen Tag oder eine Woche gesehen, bietet ein sehr gutes Beispiel hierfür. Der Nutzer muss sich darüber im Klaren sein, dass es zwei Perioden dauert, bis er die ersten verlässlichen Vorhersagen bekommt und man danach eventuell die Parameter noch anpassen muss. Außerdem sollte man bei Ereignissen, von denen man schon im Voraus weiß, dass sie etwas Außergewöhnliches darstellen, wie zum Beispiel Feiertage unter der Woche oder Backups, die nur einmal im Monat/Jahr durchgeführt werden, die Alarmierung einschränken beziehungsweise die Alarmierten darüber informieren. Beachtet man dies alles, ist man durch den Holt-Winters-Algorithmus der Zeit immer einen kleinen Schritt voraus und kann somit überprüfen, ob sich der Netzwerkverkehr so verhält, wie man es gewohnt ist und wie er sollte, oder ob er sich anomal verhält und man weitere Untersuchungen anstellen sollte, um die Gründe dafür herauszufinden. (avr)

PASCAL SCHÖTTLE

studiert Mathematical Engineering an der Uni Duisburg-Essen und hat sich mit dem Thema während eines Praktikums bei der science+computing ag befasst.

THOMAS GÖTZ

entwickelt Lösungen im Bereich der Netzwerk- und Applikationssicherheit bei der Tübinger science+computing ag.

Onlinequellen

Cricket
cricket.sourceforge.net

RRD-Tool
oss.oetiker.ch/rrdtool/



Auslaufmodell Harddisk

Andrej Mücke

Festplatten sind bislang das bevorzugte Speichermedium. Doch während ihre Kapazitäten stetig wachsen, treten andere Leistungsmerkmale wie Zugriffszeiten auf der Stelle. Anspruchsvolle Anwendungen stoßen an die Grenzen der Technik.

Das Größenwachstum der Harddisks in den letzten 20 Jahren beeindruckt: Ihr maximales Fassungsvermögen stieg von 100 MByte im Jahr 1987 auf derzeit 1000 GByte. Diese Entwicklung ist auf die stetige Erhöhung der flächenbezogenen Speicherdichte (Areal Density) zurückzuführen, aus der eine Steigerung sowohl der Datenmenge pro Spur (Linear Density) als auch der Anzahl der Spuren pro Plattenoberfläche (Track Density) resultiert. Die Gesamtoberfläche einer Festplatte nahm in diesem Zeitraum kontinuierlich ab, zum einen durch das Verkleinern der Bauform von 5,25“ über 3,5“ bis hin zu 2,5“, die in Notebooks und mittlerweile auch in Servern zum Einsatz kommen. Zum anderen verringerte sich durch die geringere Bauhöhe und aus Kostengründen die Anzahl der Scheiben. Schien eine technische Grenze erreicht, fanden sich nach kurzer Zeit neue Verfahren wie früher GMR (Giant Magneto Resistance) oder heute Perpendicular Recording, die noch mehr Informationen auf gleicher Fläche unterbringen.

Für die maximale Informationsdichte gibt es eine physikalische Grenze – zurzeit ist aber noch nicht abschätzbar, wie weit sich zukünftige Verfahren ihr nähern können.

Neben dem Speichervolumen sind die Transferraten gestiegen. Das betrifft sowohl die Geschwindigkeit der Schnittstellen als auch das reale Tempo beim physischen Schreiben/Lesen auf der Platte. Im Vergleich zu den Kapazitäten jedoch sind die Netto-Transferraten um zwei Zehnerpotenzen weniger stark gewachsen: von unter 1 MByte/s bis auf mehr als 100 MByte/s. Der Grund: Die Erhöhung der Track Density wirkt sich nicht auf die Übertragungsgeschwindigkeit aus – nur die Steigerung von Linear Density und Plattendrehzahl erhöhen sie. Zum Bremsklotz wird die geringere Zunahme der Transfergeschwindigkeit bei Vorgängen, die in Relation zur Gesamtkapazität stehen: Das Kopieren oder Prüfen der kompletten Platte dauert dadurch umso länger, je größer sie ist.

Transferraten halten nicht Schritt

Leider lassen sich die maximalen Transferraten nur bei rein sequenziellen Operationen erreichen. Bei nicht-sequenziellem Zugriff erfolgt zwangsläufig eine Neupositionierung des Kopfträgers auf die gewünschte Spur. Dieser mechanische Vorgang durchläuft zunächst die beiden etwa gleich langen Phasen der Beschleunigung und der Verzögerung des Arms. Anschließend erfolgt die Feinpositionierung auf die Zielspur. Sie dauert je nach Spurabstand und Plattenbauweise zwischen 0,3 und 25 ms. Die mittlere Zugriffszeit liegt typischerweise im Bereich von 4 bis 12 ms. In der Regel sind die Daten einer Applikation allerdings nicht über die gesamte Platte verstreut, wodurch sich die reale mittlere Zugriffszeit auf ungefähr 1 bis 4 ms reduziert.

Hinzu kommt die Zeit, bis auf der ausgewählten Datenspur der angeforderte Block den Schreib-/Lesekopf passiert. Im Mittel ist dafür eine halbe Umdrehung erforderlich. Je nach Drehzahl beansprucht diese nochmals 2 bis 6 ms. Daraus ergibt sich eine Obergrenze von 100 bis 300 I/O-Operationen pro Sekunde. Im ungünstigen Fall liest die Software pro Operation nur einen Datenblock von 4 KByte, woraus sich eine effektive Transferrate von nur noch 400 bis 1200 KByte/s ergibt. Selbst bei größeren Datenblöcken von 16 oder 32 KByte liegt die erreichbare Übertragungskapazität ganz erheblich unter der Maximalrate. Das Lesen einer kompletten 1-TByte-Disk kann auf diese Art mehrere Wochen dauern.

Grenzen der Mechanik

Bei den mechanischen Komponenten sind zu vertretbarer Kosten-Nutzen-Relation praktisch keine Leistungssteigerungen mehr zu erwarten. Eine deutliche Steigerung der Drehzahl stellt höhere Anforderungen an Antrieb und Lagerung des Plattenstapels und erfordert die Abführung der zusätzlichen Reibungswärme. Dies bedingt eine erheblich aufwendigere und damit teurere Konstruktion. Bei mehr als 10 000 Umdrehungen pro Minute ist aufgrund der größeren auf das Speichermedium wirkenden Fliehkräfte auch eine Reduzierung des Plattendurchmessers nötig – zu Lasten der Kapazität. Auf dem Desktop ist daher keine Steigerung über die zurzeit eingesetzten 7200 Umdrehungen pro Minute hinaus zu erwarten. Auch für Server lohnt es sich nicht, die derzeit maximal erhältlichen 15 000 Umdrehungen pro Minute zu steigern, da die Mehrkosten in keinem Verhältnis zu der geringen Verbesserung von Durchsatz und Latenz stehen.

Im Wesentlichen bestimmen Antrieb und Gewicht

des Kopfträgers die Geschwindigkeit der Kopfbewegungen. Diese Komponenten haben aber kaum noch – bezahlbares – Verbesserungspotenzial. Zwar lassen sich die Zugriffszeiten reduzieren, indem man den Plattendurchmesser und damit die Wegstrecke der Köpfe verringert. Allerdings würde dies selbst bei einer Halbierung der nutzbaren Oberfläche nur eine Verbesserung von circa 20 Prozent bringen. Die heutige Festplattenmechanik ist somit weitgehend ausgereizt.

Zur Leistungssteigerung integrierte man schon frühzeitig einen RAM-Cache in die Steuerungshardware. Er ist im Laufe der Zeit immer weiter gewachsen und fasst heute meistens zwischen 8 und 32 MByte. Aufgrund seiner im Verhältnis zur Gesamtkapazität sehr geringen Größe eignet sich dieser Speicher vor allem für das Prefetching. Dabei liest der Controller nach dem gewünschten Datenblock weitere Blöcke derselben Spur quasi auf Vorrat ein. Für eine spürbare Beschleunigung wahlfreier Zugriffe ist der Puffer jedoch schlicht zu klein. Außerdem speichern die Betriebssysteme Plattenzugriffe selbst zwischen, wodurch die Festplatte häufig gelesene Datenblöcke gar nicht mehr liefern muss.

Cache-Nutzung mit Risiko

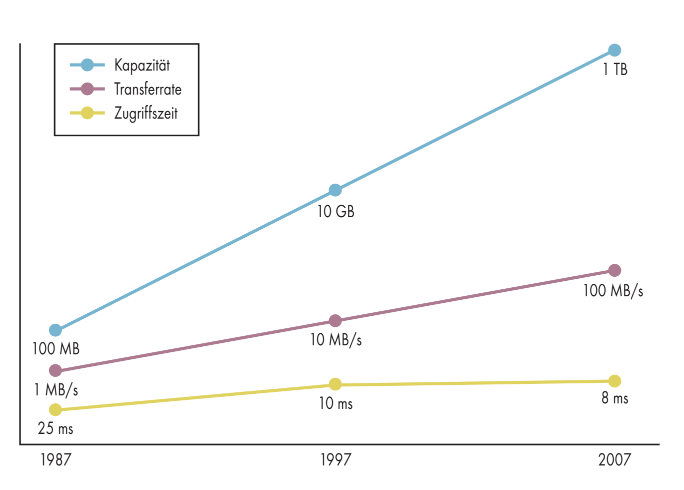
Die Nutzung des Cache für die Beschleunigung von Schreiboperationen durch verzögertes Schreiben ist riskant, da die Gefahr von Datenverlust bei Stromausfall besteht. Außerdem lässt sich oft nur eine marginale Geschwindigkeitssteigerung erzielen, da auch die Betriebssysteme das Schreiben verzögern.

In neueren Hybrid-Modellen soll Flash-Speicher die Leistung verbessern und ausfallsichere Zwischenspeicherung im Puffer das Schreiben beschleunigen. Um eine effizientere Nutzung des Caches für Lesezugriffe zu erreichen, muss allerdings das Betriebssystem die Belegung des Puffers steuern. Durch eine geeignete Gewichtung lassen sich Datenblöcke mit höherer Zugriffshäufigkeit bevorzugt dort ablegen. Die zu erwartende Beschleunigung hängt davon ab, wie hoch der Prozentsatz eingesparter wahlfreier Operationen in einer konkreten Anwendung ist.

Das Geschwindigkeitsverhalten einer Festplatte ist somit durch schnelle sequenzielle und langsame nicht-sequenzielle Zugriffen charakterisiert. Arbeitet die Software hauptsächlich mit sequenziellen Operationen (wie bei Audio/Video-Streams), spielen die Zugriffszeiten keine große Rolle. Bei Anwendungen mit vielen und stark verteilten Datenblöcken (wie Datenbanken), wirkt die Stagnation der Zugriffszeiten zunehmend als Bremsklotz: Stark steigende Datenvolumina behindern immer öfter einzelne Anwendungsbereiche spürbar.

Mehr RAM hilft etwas

Im Multi-User-Betrieb ist durch das Verteilen gleichzeitiger Zugriffe auf mehrere Platten innerhalb eines RAID-Verbunds eine gewisse Beschleunigung des Durchsatzes möglich, eine joborientierte Verarbeitung profitiert davon allerdings kaum. Software hat wenige Möglichkeiten zur Beseitigung der Performance-Engpässe: Sie kann versuchen, wahlfreie I/O-Operationen zu reduzieren. Die erfolgversprechendste Strategie dazu besteht in stärkerer Nutzung des Hauptspeichers. Dies setzt zwar eine Mindestmenge an RAM voraus, dafür kann die Applikation im Idealfall aber den gesamten Datenbestand mit wenigen sequenziellen Lesevorgängen in den Speicher laden. Alternativ lassen sich gelesene Daten über einen längeren Zeitraum im



Im Vergleich zur Kapazität haben sich Transferraten und Zugriffszeiten von Festplatten in den letzten 20 Jahren signifikant weniger verbessert.

Hauptspeicher halten, um ein Nachladen zu vermeiden und dadurch ebenfalls die Anzahl der Ein-/Ausgabeoperationen zu reduzieren.

Schwieriger ist eine Änderung der Struktur der gespeicherten Daten, die den Anteil sequenzieller Operationen erhöht. Auch das Resultat solcher Modifikationen befriedigt nicht immer. Beispielsweise liest ein Datenbanksystem bei der Vergrößerung von Datenblöcken zwar mehr Daten pro Operation, falls die Anwendung diese aber nicht kurzfristig benötigt, kann der Beschleunigungseffekt gegen Null gehen.

Alternative Solid State Disk

Als Alternative zu herkömmlichen Festplatten kommen in letzter Zeit zunehmend Solid State Disks (SSD) aus Flash-Speicher-Chips in Betracht. Da sie auf mechanische Komponenten verzichten, sind die Zugriffsraten konkurrenzlos niedrig. Mit weniger als 0,3 ms im Durchschnitt lassen sich über 3000 Random-I/Os pro Sekunde durchführen. Dies gilt bislang jedoch nur für die Lesezugriffe. Beim Schreiben vieler kleiner Datenblöcke hingegen wachsen die Zugriffszeiten stark und liegen auf dem Niveau derjenigen von Festplatten. Die Ursache für diesen Effekt sind die größeren Speicherblöcke (16 KByte) als bei Platten (512 Byte).

Bei sequenziellen Operationen erreichen die Transferzeiten noch nicht die Werte schneller Harddisks. Im Zuge der technischen Weiterentwicklung ist aber sowohl eine deutliche Steigerung der Transferraten als auch eine Verbesserung der Zugriffszeiten bei Schreiboperationen zu erwarten. SSDs verfügen außerdem über eine geringe Stoßempfindlichkeit und benötigen wenig Strom: unter einem Watt gegenüber den 5 bis 17 Watt einer Festplatte. Ein angenehmer Nebeneffekt von SSDs ist der geräuschlose Betrieb. Die Nachteile liegen derzeit vor allem im hohen Preis, den im Vergleich zu Festplatten geringeren Kapazitäten und niedrigeren Transferraten.

Fazit

Unter Performance-Gesichtspunkten ist die Festplatte ein Auslaufmodell. Für eine Weile wird sie aufgrund des guten Preis-Kapazitäts-Verhältnisses noch einen Platz haben, wo die Zugriffsgeschwindigkeit nachrangig und hohes Speichervolumen notwendig sind. Die Zukunft gehört den Flash-Speichersystemen, die mehr Potenzial zur technischen Weiterentwicklung bieten. (ck)

ANDREJ MÜCKE

ist Leiter der Datenbankentwicklung für CONZEPT 16 bei der vectorsoft AG.

Microsofts neue Server-Konsolidierung: Windows Server Hyper-V

Mit direktem Zugriff

Christian Segor

Keine Frage, Virtualisierung ist en vogue. Microsofts diesbezügliche Angebote sind aber zurzeit für den professionellen Einsatz kaum geeignet. Zusammen mit dem Windows Server 2008 soll sich das ändern.

Wer als CIO etwas auf sich hält, hat sich, um über Virtualisierung mitreden zu können, längst mit dem notwendigen Wissen aus der Fachpresse oder zumindest der ein oder anderen Buzzword-Publikation versorgt. Die Vorteile virtualisierter IT-Systeme sind oft beschrieben und liegen zudem klar auf der Hand.

In einer sauber entworfenen IT-Infrastruktur ist die Aufteilung verschiedener Aufgaben auf unterschiedliche Server durchaus gewollt. Es mag ja noch angehen, netzlastige Dienste wie DNS, WINS oder DHCP auf einer Maschine zusammenzupacken, aber viele Administratoren verspüren berechtigterweise eine mehr oder weniger deutliche Nausea bei dem Gedanken, den MS SQL Server und Exchange auf demselben Server zu installieren oder den Intranet-IIS mit einem DC zu verheiraten.

Abhilfe kann eine konsequente Verwendung virtualisierter Hardware schaffen. Auf einem physischen System laufen mehrere logische Server-Instanzen, die Ressourcen werden besser genutzt, der beschriebene modularisierte Ansatz bleibt aber trotzdem bestehen. „Ein Service pro Server“ kann man mittels Virtualisierung überhaupt erst realisieren.

Umso erstaunlicher ist es, dass Virtualisierungsprodukte bisher keine sonderlich tiefe Marktdurchdringung erreichen konnten. Klar, jeder betreibt eine Testumgebung auf VMware & Co.; in der Produktion hingegen verlassen sich viele ITler jedoch lieber auf echtes Blech. Der Anteil virtualisierter Server schwankt je nach Marktforschungsinstitut, bewegt sich allerdings konstant um die 5 %-Marke herum. Analysten schätzen, dass sich dieser Anteil in den nächsten drei Jahren zwar erhöhen, aber immer noch deutlich unter 20 % bleiben wird.

In diesen aufstrebenden Markt hinein platziert Microsoft seine neue Virtualisierungstechnologie namens Windows Server Hyper-V – die Bestandteil des Windows Server 2008 sein wird. Hyper-V ist nicht das erste Virtualisierungsprodukt aus Redmond (man denke an Virtual PC oder Virtual Server), wohl aber das erste, das wirklich für den professionellen RZ-Betrieb geeignet zu sein scheint.

Zugriff am Host-System vorbei

Es gibt grob gesagt zwei Ansätze für die Virtualisierung von Betriebssystemen, die sich vor allem durch die Position des Virtual Machine Managers (VMM) im OS-Architekturmodell unterscheiden, also der Komponente, die die tatsächlichen Systemressourcen virtualisiert und den Gastsystemen zur Verfügung stellt. Im Fall der Hosted Virtualization (hierzu zählen Virtual PC und Virtual Server, aber auch die Konkurrenzprodukte VMware Workstation und VMware Server) läuft der VMM entweder irgendwo neben dem gastgebenden Betriebssystem im Kernel oder oberhalb als User-Prozess. Beim Zugriff auf Ressourcen sind diese VMs somit auf das Wohlwollen des Host-OS angewiesen, das den VMM im Zweifelsfall als einen Prozess von vielen betrachtet.

Hyper-V hingegen bringt einen VMM mit, der – ähnlich dem Xen-Ansatz – direkt auf der Hardware läuft, was als Native Virtualization bezeichnet wird. Der ESX-Server von VMware ist eine leichte Abwandlung dieses zweiten Modells. Einen Grundlagenartikel zur Virtualisierung von Betriebssystemen hat *iX* in der Januar-Ausgabe veröffentlicht [1].

Microsoft selbst bezeichnet diesen speziellen Typus eines VMM als „Hypervisor“, anderen Ortes werden die Begriffe „VMM“ und „Hypervisor“ synonym benutzt. Da der VMM hier über einen direkten und exklusiven Zugriff auf die Hardware verfügt, sind VMs nicht mehr auf die Kooperation des Host-OS angewiesen, vielmehr läuft das Host-OS mehr oder minder parallel und

gleichberechtigt zu den verschiedenen Gast-OS (siehe Abbildung 1).

Ob man in diesem Fall überhaupt noch von Host-OS reden kann, sei dahingestellt – Hyper-V unterscheidet dann auch zwischen einer Parent Partition, die gewissermaßen die Rolle des Host-OS übernimmt und in der diverse Kontrollprozesse laufen, und den Child Partitions, die die eigentlichen VMs darstellen. Läuft in einer VM ein Betriebssystem, das Hypervisor-aware ist, kann diese VM direkt auf die von Hyper-V zur Verfügung gestellten virtuellen Ressourcen zugreifen. Die virtuellen Geräte werden von sogenannten Virtualization Service Providers (VSP) in der Parent Partition angeboten und von Virtualization Service Clients (VSC) in jeder Child Partition benutzt. VSP und VSC kommunizieren über eine Art Software-Bus namens VMBus (siehe Abbildung 2).

Momentan sind nur Windows Server 2003 und 2008 Hypervisor-aware, für alle anderen Gast-Betriebssysteme kommt eine Emulation zum Einsatz, die zwar die Performance beeinträchtigt, die grundsätzliche Funktionen aber beinhaltet, indem sie VSC und VMBus zum Gast-OS hin kapselt.

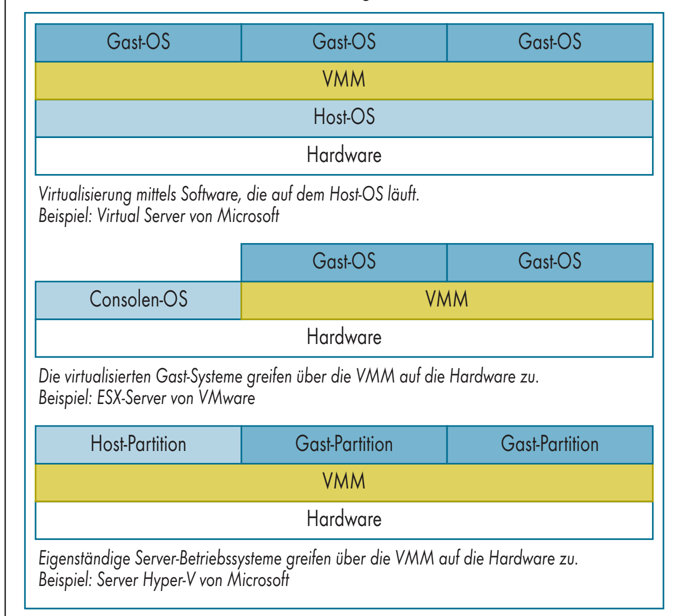
Bei den Gast-OS kann es sich sowohl um 32-Bit- als auch um 64-Bit-Systeme handeln, im letzten Fall können mehr als 4 GByte virtueller Hauptspeicher zur Verfügung gestellt werden. Bei Bedarf kann der Administrator einer VM direkten Zugriff auf eine physikalische Festplatte einrichten, andernfalls speichern VMs ihre Daten auf virtuellen Platten im VHD-Format [2]. Hyper-V unterstützt den Volume Shadow Service für Datensicherungen und ermöglicht so dem Administrator, Snapshots virtueller Maschinen anzufertigen; sinnvoll etwa vor dem Einspielen von Servicepacks und Hotfixes.

In virtualisierten Umgebungen steigen die Anforderungen an die Verfügbarkeit der gemeinsam genutzten Hardware – durchaus verständlich, kann doch ein Hardware-Problem oder auch nur eine notwendige Wartung unter Umständen den Ausfall zahlreicher virtueller Server bedeuten. Beim professionellen Einsatz virtueller Server ist somit eine Konstellation vonnöten, die es dem Administrator erlaubt, VMs ohne nennenswerte Ausfallzeiten von einem System auf ein anderes zu verlagern.

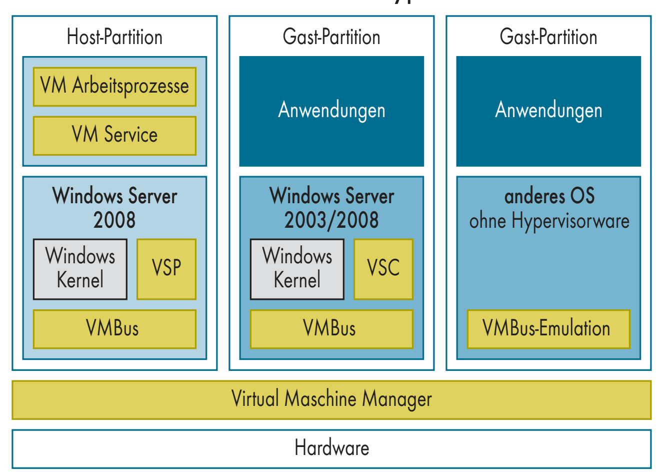
Virtuelle Maschinen im Cluster

Zu diesem Zweck unterstützt Hyper-V die Cluster-Funktionen des Windows Server 2008. Neben einer geplanten Verlagerung der VMs bieten sie einen automatischen Failover, sodass unerwartete Hardware-Ausfälle (die es bei sinnvoll überwachter Qualitätshardware eigentlich kaum mehr geben dürfte) mit nur geringen Ruckeleien und ohne Eingriff des Administrators gemeistert werden. Ferner unterstützt der neue Server den Aufbau von Geo-Clustern mit einfachen Mitteln (die Verteilung der Cluster-Knoten auf mehrere Standorte), was für Hochverfügbarkeits-Jünger ziemlich attraktiv ist.

Hyper-V lässt sich als zusätzliche Serverrolle auf jeder 64-Bit-Variante des neuen Servers installieren – dabei schiebt sich der VMM gewissermaßen zwischen OS und Hardware. Eine Hyper-V-Installation auf der Standard Edition ist wohl nur für Testumgebungen sinnvoll. Für den Produktiveinsatz sollte der Administrator mindestens zur Enterprise Edition greifen, nicht zuletzt aus Lizenzgründen, sind doch in dieser Lizenz zusätzlich vier für virtualisierte Server enthalten. Die Datacenter Edition beinhaltet sogar unlimitierte virtuelle Nutzungsrechte.



Die Hosted Virtualization ist für den professionellen Einsatz kaum geeignet, da die Gastsysteme mit allen ihren Funktionen auf das Betriebssystem angewiesen sind, unter dem der VMM läuft (Abb. 1).



Alle Betriebssysteme, auch das in der Host-Partion, greifen direkt über den VMM auf die Hardware zu (Abb. 2).

Es ist erfreulich, dass sich Hyper-V auch auf dem sogenannten Server Core installieren lässt. Dabei handelt es sich um eine stark reduzierte Variante vom 2008-Server, die ohne GUI und sonstigen Schnickschnack daherkommt [3]. Diese Reduktion soll die Stabilität erhöhen und verringert gleichzeitig die Fläche für potenzielle Angriffe. Hyper-V auf einem Server Core eignet sich beispielsweise ganz vorzüglich für die Realisierung von virtualisierten Zweigstellen-Konzepten und Ähnlichem.

Voraussetzungen und Verfügbarkeit

Neben der 64-Bit-Fähigkeit der Hardware benötigt Hyper-V einen Prozessor, der VT von Intel [4] oder AMD-V [5] unterstützt, sowie hardwarebasierte „Data Execution Prevention“ [6]. Eine Beta von Hyper-V soll gleichzeitig mit der finalen Version von Windows Server 2008 erscheinen, die Release ist spätestens 180 Tage danach fällig. Weitere drei Monate später wollen die Redmonder den Hyper-V-Server auf den Markt bringen, eine Art Stand-alone-Variante von Hyper-V, die ohne Server 2008 betrieben werden kann.

Momentan ist Hyper-V als Customer Technology Preview (CTP) für den RC0 der 2008-Version verfügbar. Der interessierte Administrator kann diese CTP mittels Windows Update direkt auf eine RC0x64-Installation herunterladen. Die Installation auf Server Core ist mit der CTP nicht möglich. (WM)

CHRISTIAN SEGOR

ist Technologieberater bei der Glück & Kanja Consulting AG in Stuttgart.

Literatur

[1] Fred Hantelmann; Xenologie; Aufbau virtueller Maschinen; *iX* 1/07, S. 130

[2] Microsoft Corp.; Virtual Hard Disk Image Format Specification; Whitepaper, Oktober 2006

[3] Nils Kaczenski; Windows ohne Windows; Microsofts Server Core: Vorbote des modularen Betriebssystems; *iX* 5/07, S. 106

[4] Intel Corp.; Enhanced Virtualization on Intel Architecture-based Servers; Whitepaper, Februar 2006

[5] Advanced Micro Devices Inc.; Putting Server Virtualization to Work; Whitepaper, Januar 2007

[6] Microsoft Corp.; A detailed description of the Data Execution Prevention (DEP) feature in Windows XP Service Pack 2, Windows XP Tablet PC Edition 2005, and Windows Server 2003; KB-Artikel 875352, September 2006

Google Guice: Framework für Dependency Injection

Codeabhängig

Tobias Lütticke, Christian Meder

Google Earth und Maps sind Beispiele für die Innovationsfreude des Suchmaschinengiganten. Weitere Projekte haben sich in Googles hausinterner Entwicklung bewährt. Ein junger Spross dieser Familie ist das Open-Source-Framework Guice, eine schlanke Alternative zu Spring.

Google Guice, ein Open-Source-Framework für die Dependency Injection (DI), entstand während des Ausbaus der Internet-Werbeanwendung AdWords und sollte dort Probleme bei der Team-Skalierung beheben, die in Projekten mit mehreren Hundert Entwicklern auftreten. Trotz anerkannter DI-Frameworks wie dem populären Spring [1] wollte Google eine eigene, besonders leichtgewichtige Variante für die interne Softwareentwicklung einsetzen (siehe Kasten „Lose Beziehung").

Generics und Annotations aus Java5 sind wesentlicher Bestandteil des Framework [2]. Zudem unterstützt Guice eigene Scopes (siehe Glossar), aspektorientierte Programmierung (AOP) sowie die Spring-Integration. Die Fähigkeit zum Injizieren statischer Attribute sowie das Management zirkulärer Referenzen runden das Bild ab.

Als Beispiel dient hier eine Anwendung zur Urlaubsplanung, die Hotels bucht (Abbildung 1). Das Planer-Objekt (Aufrufer) spricht ein Interface *Booking* an, hinter dem sich konkrete Implementierungen (Zielobjekte) verbergen. Um während des Entwickelns nicht fortwährend das echte Zielsystem abfragen zu müssen, kommt hier ein sogenanntes Mock-Objekt (Dummy) zum Einsatz. Für den Produktivbetrieb soll der Entwickler einfach umschalten können.

Das Framework soll besonders schlank wirken und den „Boilerplate Code" massiv reduzieren. Statt Fleißarbeit zu leisten, soll sich der Entwickler mit dem Wesentlichen beschäftigen. Dazu konzentriert sich Guice ausschließlich auf eine performante Umsetzung der DI und ist entsprechend abgespeckt, sowohl was die Größe der JAR-Files als auch den Speicherverbrauch angeht. Die Abhängigkeiten verwaltet Guice nicht in zentralen XML-Dateien, sondern hinterlegt sie im Java-Code.

Konflikte lösen mit Java

Erfahrungen haben gezeigt, dass die Entwicklung nicht skaliert, wenn zu viele Beteiligte um den Zugriff auf Kernkomponenten konkurrieren. Hässliche Konflikte sind bei der Arbeit mit gemeinsam genutzten XML-Dateien unvermeidbar. Das Bearbeiten von Java-Dateien ist zudem einfacher und weniger fehlerträchtig als das von großen XML-Dateien. Das Refactoring wird dadurch spürbar erleichtert und durch Tools besser unterstützt.

Den Praxistest hat Guice in AdWords bereits bestanden. Außerdem ist eine frühe Version Bestandteil des bekannten Web-Frameworks Struts2, in dem sie das zentrale Element der Plug-in-Architektur stellt. Guice kann in allen Java-Anwendungen nutzbringende Dienste verrichten. Besonders profitieren große Projekte und solche, die sich ausschließlich auf eine effiziente Dependency Injection konzentrieren. Hier kann Guice seine Stärken wie Schlankheit und XML-freies Wiring ausspielen (siehe Kasten „Gut verdrahtet").

Als Beispiel dient wieder die Anwendung zur Urlaubsplanung. Guices Aufgabe ist es, dem Planer das gewünschte konkrete Booking-Objekt zuzuordnen, auf das er über das Interface zugreift. Im betrachteten Fall soll dieses das Objekt *MockBooking* sein. Dazu muss geklärt sein, was wo injiziert wird. Den ersten Teil der Frage beantwortet ein sogenanntes Modul mit einem „Binder", die gemeinsam das Interface der Implementierung zuweisen. Der folgende Code-Schnipsel bildet ein solches Modul und bindet mithilfe der *bind()*-Methode das Interface *Booking* an die konkrete Klasse *MockBooking*.

```
public class PlanerModule↵
                        implements Module {
  public void configure(Binder binder) {
    binder.bind(Booking.class).to↵
                        (MockBooking.class);
  }
}
```

Um das „wo" kümmert sich eine *@Inject*-Annotation im *Planer*-Objekt:

iX-TRACT

- Google Guice ist ein leichtgewichtiges Framework für das Programmierkonzept „Dependency Injection".
- Im Gegensatz zu alternativen Angeboten beschränkt sich Guice auf ein Thema; Spring beispielsweise bietet einen erheblich größeren Funktionsumfang.
- Außerhalb von Google gibt es bislang kaum Guice-Anwendungen. Durch die Marktmacht des Suchmaschinenanbieters könnte sich die Situation jedoch künftig anders darstellen.

Glossar

Boilerplate Code: Sich wiederholender, aber erforderlicher Schema-F-Code, der keine Geschäftslogik abbildet.

Injector: Komponente, die eine Abhängigkeit erfüllt, indem sie das gewünschte Objekt injiziert.

Modul: Container für Bindings, der festlegt, was zu injizieren ist.

Binding: Verknüpfung zwischen einem Interface und seiner Implementierung.

Scope: Gültigkeitsbereich eines injizierten Objekts (zum Beispiel Request, Session, Singleton).

(Custom) Provider: Erzeugt die zu injizierenden Objekte. In der Variante „Custom“ zur individuellen Anbindung applikationsfremder Komponenten.

Bootstrapping: Initialisierung von Guice durch Injektion des Wurzelobjekts, um das Auto-Wiring anzustoßen. Das vermeidet, später für jedes Injizieren den Injector erneut abfragen zu müssen.

Annotations: Meta-Informationen, die der Entwickler im Quellcode der Java-Klassen mit vorangestelltem „@“-Zeichen hinterlegt.

Generics: Mit Typparametern parametrisierbare Methoden und Klassen.

Gut verdrahtet

Wiring bezeichnet die Art und Weise, wie das Framework Abhängigkeiten zwischen Objekten verwaltet. Es existieren zwei grundlegende Ausprägungen: die explizite Darstellung in XML beziehungsweise Java und das automatische Auswerten durch das Framework zur Laufzeit (Auto-Wiring).

Auswerten lassen sich beide Varianten. Im Falle des Auto-Wiring versucht das Framework, die Beziehungen zwischen den Komponenten selbst zu ergründen. Das erzeugt allerdings mit zunehmender Komplexität der Anwendung ein Problem: Die Performance wird immer schlechter.

Lose Beziehung

Dependency Injection (DI) ist ein populäres Konzept in der Softwareentwicklung. Zwei der primären Ziele: Wiederverwendbarkeit und Wartung von Software durch lose Kopplung von Komponenten zu vereinfachen sowie das Testen zu erleichtern.

Wenn ein Objekt *A* eine Referenz auf ein Objekt *B* hält, kennt es in der Regel dessen Implementierung. Das DI-Konzept sieht vor, dass die Abhängigkeiten zwischen dem Aufrufer *A* und von ihm angesprochenen Zielen *B* nicht mehr in *A* hinterlegt werden. Welche Realisierung des Zielobjekts *B* konkret anzusprechen ist, wird dem Aufrufer erst bei Bedarf mitgeteilt. Die Beziehung zu dem konkreten *B* wird dazu in *A* „injiziert“. Als Aufrufer kennt *A* also das zu verwendende Zielobjekt bis zur endgültigen Injektion nicht.

DI ist eine Anwendung des Paradigmas Inversion of Control (IoC), auch Hollywood Principle genannt: Don’t call us, we’ll call you. Auf diese Weise erlaubt es Dependency Injection, verschiedene Nutzungsszenarien zu verwenden, ohne den Quellcode des Aufrufers dafür anpassen zu müssen, etwa wenn neue Realisierungen der Zielobjekte hinzukommen. Gängiges Beispiel für ein solches weiteres Ziel ist ein Service als vereinfachte Variante des ursprüngliches Dienstes für Testzwecke. Wenn der eigentliche Dienst viele Ressourcen braucht und lange läuft oder überhaupt nur in der Produktivumgebung zur Verfügung steht, erlaubt die Ausprägung als Mock-Objekt (Dummy) ein (schnelleres) Testen.

```
public class Planer {
  private final Booking booking;
  @Inject
  public Planer(Booking booking) {
    this.booking = booking;
  }
}
```

Die *@Inject*-Annotation gibt an, dass mithilfe des Konstruktors ein Objekt vom Typ *Booking* zu injizieren ist. Das Framework beherrscht Konstruktor-Injection (wie im Beispiel) sowie Method und Field Injection. Das bedeutet, dass Guice auch beliebige annotierte Methoden nutzen oder direkt in ein Attribut injizieren kann:

```
@Inject String name;
```

Injizierbare Elemente sind primitive Typen wie *int* oder *char*, außerdem Enums und ganz allgemein Instanzen beliebiger Klassen.

Das Testszenario lässt sich weiter vereinfachen, indem man auf das PlanerModule mit seinem Binding verzichtet und im Interface mithilfe der Annotation *@ImplementedBy* ein Default-Binding angibt:

```
@ImplementedBy(MockBooking.class)
public Interface Booking {
// Methoden
}
```

Hat der Entwickler die Abhängigkeiten modelliert, kann Guice seine Tätigkeit beginnen. Es unterscheidet Initialisierung und Laufzeit. Beim Start der Applikation beginnt das Bootstrapping mit dem Injizieren eines Wurzelobjekts. Den Rest erledigt Guice, indem es sich durch den Abhängigkeitengraphen hangelt und rekursiv alle weiteren Injektionen vornimmt. Dabei findet eine Validierung statt, die auf Lücken oder Fehler hinweist. Die Verwendung ist so einfach wie in diesem Beispiel:

```
public class Urlaubsplanung {
  public static void main(String[]↵
                              args){
    Injector injector = Guice.create↵
                          Injector();
    Planer planer = injector.↵
                getInstance (Planer.class);
    // Geschäftslogik
  }
}
```

Teil eines jeden Bindings ist ein Provider, der Instanzen der registrierten Klasse bereitstellt. In gewissen Fällen kann es sinnvoll oder unumgänglich sein, Objekte selbst zu erzeugen und dies nicht dem internen Standard-Provider zu überlassen. Setzt man eine fremde Bibliothek ein, so lässt sich dort zum Beispiel keine @Inject-Annotation eintragen. Mithilfe eines Custom Providers ist es aber trotzdem möglich, das Binding vorzunehmen. Custom Provider erlauben auch die Integration mit über JNDI oder JMX gelieferten Objekten.

Keine Angst vor Spritzen

Schlüsselelement in der Guice-Struktur ist der Injector, der die Bindings verwaltet (Abbildung 2). Letztere verwenden eine Annotation am Injektionsziel und beziehen sich auf einen Typ, den es zu injizieren gilt. Der Provider erzeugt die Instanzen des Typs. Der Scope eines Bindings ist eine optionale Angabe und steuert das Wiederverwenden erzeugter und injizierter Objekte. Für jede Injektion legt Guice im Standardfall eine neue Instanz des betroffenen Objekts an. Alternative Scopes sind „Singleton“, „Request“ und „Session“.

Über die beschriebenen grundlegenden Mechanismen hinaus kennt Guice etliche weiterführende Möglichkeiten. Dazu gehört die Definition eigener Annotationen, die beispielsweise eine zusätzliche Zuordnung von Flug- und Mietwagenbuchung erlaubt.

Wer Dependency Injection oder Inversion of Control hört, denkt in der Regel zuerst an Spring. Nicht zu Unrecht, handelt es sich doch um den De-facto-Standard, der eine aktive Community hinter sich versammelt hat. Wie schon erwähnt, sieht Google trotz dessen Existenz Bedarf für eine eigene Lösung. Auch Alternativen wie jBoss Seam,

Apache HiveMind oder PicoContainer hielten das Unternehmen nicht von der Eigenentwicklung ab.

Beide Frameworks – Spring und Guice – gehören zu den Open-Source-Erzeugnissen und stehen unter der Apache-2.0-Lizenz. Annotations und Generics als Bestandteile von Guice erlauben die Nutzung erst ab Java 5. Für Spring gilt das nicht, es lässt sich selbst mit dem JDK 1.3 einsetzen und bietet deutlich mehr Funktionen als Guice. Während sich Letzteres auf DI konzentriert, unterstützt Spring zudem Transaktionen, Persistenz und besitzt ein eigenes Web-Framework. Subprojekte existieren für Konfiguration, Security, Batch-Jobs und einige mehr.

Die Definition von Abhängigkeiten zwischen Objekten bildet das Rückgrat eines DI-Framework. Wie die Abhängigkeiten hinterlegt und ausgewertet werden (Wiring der Objekte) ist eine kennzeichnende Eigenschaft. Hier zeigt sich ein weiterer wesentlicher Unterschied zwischen Spring und Guice. Spring erlaubt sowohl die explizite Darstellung als auch das Auto-Wiring. Guice verfolgt einen anderen Ansatz: Es setzt zwar auf eine explizite Darstellung, umgeht aber das geschwätzige XML-Format, indem es die Zusammenhänge über Java-Annotationen im Quellcode unterbringt. Dieses Vorgehen lässt sich als Mischung zwischen der ausführlichen, aber wartungsintensiven expliziten Darstellung und dem Auto-Wiring auffassen. Spring erlaubt zwar mit *JavaConfig* ebenfalls das Anlegen von Abhängigkeiten im Java-Code, offiziell befindet sich dieses Subprojekt allerdings noch im Alpha-Zustand.

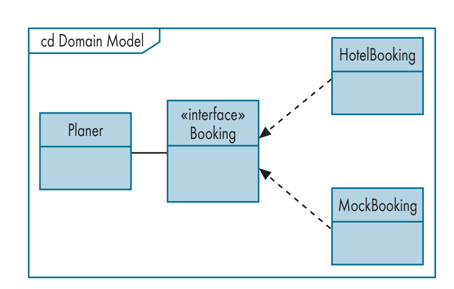
Klein, schnell, übersichtlich

Seine Performance-Vorteile gegenüber Spring spielt Guice sowohl beim Start der Anwendung als auch zur Laufzeit aus, wenn es um das Erzeugen der angeforderten Objekte geht. Den Vorsprung bei der Anlage der Objekte erzielt Guice durch die Java5-Fähigkeiten zur Nebenläufigkeit. Außerdem verzichtet es auf das Erzeugen der Proxy-Objekte, die Spring einsetzt. Wie relevant das in der Praxis ist, hängt jedoch vom Charakter der Applikation ab.

Neben den technischen Eigenschaften sowie dem Funktionsumfang zählen bei der Wahl eines Werkzeugs Aspekte wie seine Komplexität. Google sieht Guice in Bezug auf Einfachheit, Übersichtlichkeit, Wartbarkeit, Erlernbarkeit und Performance gegenüber Spring klar im Vorteil.

Insgesamt haben die zahlreichen Vergleiche der beiden Frameworks in den entsprechenden Kreisen Aufregung erzeugt. Letztlich läuft es jedoch nicht auf eine Entweder-oder-Entscheidung hinaus, da zwar beide DI-Container sind, jedoch einen unterschiedlichen Fokus setzen: Guice hinterlässt einen kleinen „Footprint“, während sich Spring als umfassende und mächtige Lösung anbietet. Ein Gutteil ihrer XML-Komplexität resultiert zudem aus Funktionen außerhalb von DI. Da Guice diese Bereiche nicht abdeckt, entsteht keine echte Rivalität. Die Koexistenz beider Frameworks ist denkbar, denn Guice erlaubt das Einbinden von Spring Beans.

Der Plan für Guice 2 enthält eine ganze Reihe von Erweiterungen, unter anderem Construction Listener und eine Introspection API. Die Listener bieten Einstiegspunkte bei der Objekterzeugung zum Einhaken eigener Logik. Die API-Erweiterung soll Interna des *Injector*-Objekts offenlegen und vollständigen Zugang zum Abhängigkeitsgraphen erlauben. Das würde beispielsweise Visualisierungswerkzeugen einen Ansatzpunkt bieten. Ferner nennt die Roadmap Provider-Methoden, mit denen sich Modulklassen einfacher gestalten lassen sollen. Custom Provider könnten damit ein flexibleres Binding durchführen, das zum Beispiel nur in einem bestimmten Scope wirksam wird.

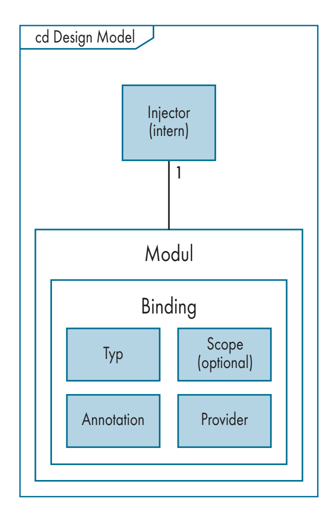


Der Planer greift über ein *Booking*-Interface auf die konkreten Klassen *HotelBooking* und *MockBooking* zu (Abb. 1).

Keine Eintagsfliege

Guice ist noch so neu, dass außerhalb von Google wenige produktive Anwendungen existieren. Eine Ausnahme bildet das erwähnte Web-Framework Struts2. Der Google-Hintergrund und der Einsatz in AdWords werden aber dafür sorgen, dass Guice nicht als Eintagsfliege endet. Weiteres Indiz dafür sind die mittlerweile existenten Third-Party-Module. Dazu gehören etwa Komponenten, die das Zusammenspiel von Guice mit Hibernate oder dem Ajax-Framework DWR ermöglichen.

Auf der Sollseite schlägt sich die noch recht übersichtliche Projektdokumentation nieder. Die zunehmend erscheinenden Erfahrungsberichte und Anleitungen mildern dieses Manko etwas ab. Vom architektonischen Standpunkt aus ist außerdem zu bedenken, dass die proprietären Annotations die Wiederverwendbarkeit von Klassen in Nicht-Guice-Projekten erschweren. Guice ist relativ leicht zu erlernen und seine Einführung einfach – auch in bestehende Projekte. Dass sich das Framework schrittweise installieren lässt, ist ein weiterer Vorteil. (jd)



Abhängigkeiten verwaltet Guice in Form von Bindings, die aus zu erzeugendem Objekttyp, Annotation und Scope bestehen. Der Provider liefert das Objekt an den Injector (Abb. 2).

TOBIAS LÜTTICKE

arbeitet als Solution Architect in einem Ministerium in Wellington, Neuseeland.

CHRISTIAN MEDER

ist Seniorarchitekt bei der inovex GmbH in Pforzheim.

Literatur

[1] Tobias Himstedt; Java-Entwicklung; Frühlingsgefühle: Spring-Framework: J2EE-Module; *iX* 7/2005, S. 132

[2] Markus Eisele; Java-Spezifikation; Nächste Runde; Einfacher programmieren mit Java Enterprise Edition 5; *iX* 7/2006, S. 136

Onlinequellen

Googles Entwicklerseiten
code.google.com

Guice
code.google.com/p/google-guice/

Dependency Injection
martinfowler.com/articles/injection.html

AdWords
adwords.google.com

Spring
www.springframework.org



SPARQL – Anfragesprache für RDF-Daten

Funkenschlag

Stefan Mintert, Bastian Spanneberg

RDF heißt die Basis des vom World Wide Web Consortium propagierten Semantic Web. Mit Metadaten angereicherte Dokumente sollen Maschinen „Verstehen“ ermöglichen. Für die Suche in solch semantisch angereicherten Webdokumenten hat das Konsortium als Anfragesprache SPARQL vorgesehen.

Mit dem Resource Description Framework (RDF) hat das World Wide Web Consortium (W3C) vor drei Jahren seine Empfehlung zur Repräsentation von Metadaten in einem gerichteten, mit einem Label versehenen Graphen veröffentlicht. RDF ist nach XML die zweite wesentliche technische Grundlage des W3C für das Semantic Web. Mit SPARQL – einem aus „SPARQL Protocol and RDF Query Language“ gewonnenen Kunstwort, das wie das englische *sparkle* (deutsch: funkeln, glitzern) zu sprechen ist – steht das Konsortium nun vor der Verabschiedung einer für RDF geeigneten Query-Sprache.

Queries auf Daten auszuführen ist offensichtlich nur dann eine spannende Angelegenheit, wenn schon Daten vorliegen. Nun hat sich das Semantic Web in der öffentlichen Wahrnehmung nicht als das große Thema eines neuen Web erwiesen, als das es viele vor einigen Jahren gesehen haben. Stattdessen steht im Mittelpunkt des sogenannten Web 2.0 der Mensch, und das „neue“ Web heißt oft „Mitmach-Web“. Wer jedoch genau hinsieht, findet an vielen Stellen des Web 2.0 Ideen des semantischen. Aus diesem Grund sollen Web-2.0-Daten hier als Grundlage dienen.

Die größeren der folgenden Codebeispiele verwenden RDF-Daten aus einem Semantic Media Wiki (SMW). Dieser Artikel lässt sich damit als Fortsetzung des Artikels aus der November-*iX* 2007 zu ebendiesem Thema lesen (siehe [1]). Wer RDF-Daten ganz anderer Natur besitzt, etwa die Daten des firmeninternen Wissensspeichers, braucht diese Motivation nicht. Um eine gute Nachvollziehbarkeit des folgenden Textes für alle Leser zu gewährleisten, braucht es noch ein geeignetes öffentliches Semantic Media Wiki. Wikipedia ist zwar zweifellos die größte Media-Wiki-Site, läuft aber „nur“ auf der ursprünglichen Software. Im Gegensatz dazu liegt Ontoworld.org die um Semantic-Web-Funktionen aufgepeppten Wiki-Software zugrunde. Eine Einführung in die Funktionen ist in der oben genannten *iX* zu finden. Ein RDF-Export der in der Ontoworld erfassten Metadaten steht unter ontoworld.org/RDF/ zur Verfügung. Der aktuelle Umfang liegt bei knapp 4,2 MByte – genug Masse, um mit SPARQL zu spielen.

Neben Daten und einer Anfragesprache ist selbstverständlich Software für den Datenspeicher und die Query Engine erforderlich. Für beides gibt es Open-Source-Alternativen, wie der Kasten „Onlinequellen“ zeigt. Einen Crashkurs für die hier verwendete Software – Jena, ARQ und Joseki – enthält der Kasten „Mit SPARQL arbeiten“.

Derzeit liegen drei Spezifikationen des W3C im Status „Proposed Recommendation“ vor (= nur noch einen Schritt von der Verabschiedung entfernt). Es handelt sich zunächst um die Query-Sprache selbst. Dazu gesellt sich mit „SPARQL Protocol for RDF“ eine Beschreibung, wie Queries und deren Antworten über HTTP und SOAP übermittelt werden können. Der dritte Text trägt den Namen „SPARQL Query Results XML Format“ und beschreibt den Aufbau von Query-Antworten in XML. Außer der Beschreibung steht ein Relax-NG- und ein daraus erzeugtes XML-Schema zur Verfügung.

Kein SPARQL ohne RDF

Vor dem Einstieg in die Anfragesprache müssen wenigstens grundlegende Kenntnisse über RDF vorhanden sein. Das lässt sich schnell erledigen: RDF-Ausdrücke haben den Aufbau „Subjekt Prädikat Objekt“, es handelt sich daher um Tripel. Das Subjekt macht durch das Prädikat eine Aussage über das Objekt. Idealer-

iX-TRACT

- SPARQL ist eine Query-Sprache für RDF-Daten und entspringt der Semantic-Web-Arbeit des World Wide Web Consortium (W3C).
- Ähnlich SQL für relationale Datenbanken soll SPARQL zur Standard-Anfragesprache des Semantic Web werden.
- Es existieren mehrere leistungsfähige Open-Source-Implementierungen.

weise sind alle drei Dinge eindeutig über einen URI zu identifizieren. Der einführende Text des W3C (RDF Primer, [a]) nennt folgendes Tripel in englischer Sprache: ***http://www.example.org/index.html*** *has a* ***creator*** *whose value is* ***John Smith***. In RDF lässt sich das wie folgt ausdrücken:

* Subjekt: http://www.example.org/index.html
* Prädikat: http://purl.org/dc/elements/1.1/creator
* Objekt: http://www.example.org/staffid/85740

Die nicht-formale Notation ist insofern korrekt, als RDF keine bestimmte Syntax verlangt. Eine weitverbreitete Syntax benutzt XML; die Aufzählung könnte man deshalb ebenso gut in XML und viele weitere Notationen übersetzen.

Subjekt und Prädikat sind in diesem Beispiel mit einem willkürlich gewählten, aber als eindeutig angenommenen URI mit der Angabe *example.org* bezeichnet. Der dritte URI legt den Verdacht nahe, dass John Smith ein Mitarbeiter einer fiktiven Organisation namens example.org ist. Formal abgesichert ist diese Aussage nicht; das ist auch nicht notwendig. Im Gegensatz dazu ist das Prädikat durch den Dublin Core definiert, und der URI ist bei purl.org registriert.

Wie oben erwähnt, handelt es sich bei RDF-Strukturen um gerichtete, mit einem Label versehene Graphen. Man erhält diese Repräsentation, wenn man Subjekt und Objekt als Knoten betrachtet, zwischen denen eine vom Subjekt ausgehende, gerichtete Kante, die das Label des Prädikats trägt, verläuft. Eine bildliche Darstellung zeigt die dem RDF Primer entnommene Abbildung 1.

Es versteht sich von selbst, dass die Strukturen durchaus komplexere Formen annehmen können. Dabei ist es beispielsweise möglich, mehrere Aussagen über ein Subjekt zu machen. Der RDF Primer fährt im oben genannten Kontext fort:

http://www.example.org/index.html has a creation-date whose value is August 16, 1999
http://www.example.org/index.html has a language whose value is English

Die grafische Darstellung ändert sich dementsprechend, wie Abbildung 2 zeigt.

Aller Anfang ist einfach

Schon auf dieser kleinen Datenbasis lassen sich SPARQL-Queries formulieren. Eine grundlegende Anfrage ist die per *SELECT-WHERE*. Dafür gewinnt SPARQL sicher keinen Preis in Originalität, der Einstieg ist durch diese durchaus vertraute Klausel aber einfach. Die Query in Listing 1 ermittelt den Urheber der Ressource *http://www.example.org/index.html*.

Es ist unschwer zu erraten, dass es sich bei *?urheber* um eine Variable handelt. Die Query instanziiert die Variable mit allen Knoten, die über die Relation *http://purl.org/dc/elements/1.1/creator* mit dem genannten Subjekt verbunden sind. Im Beispielcode der obigen Daten ist die Treffermenge einelementig: *http://www.example.org/staffid/85740*. In der XML-Notation für SPARQL-Ergebnisse sieht das aus wie der untere Teil von

Listing 1: Simple Anfrage/Ergebnis

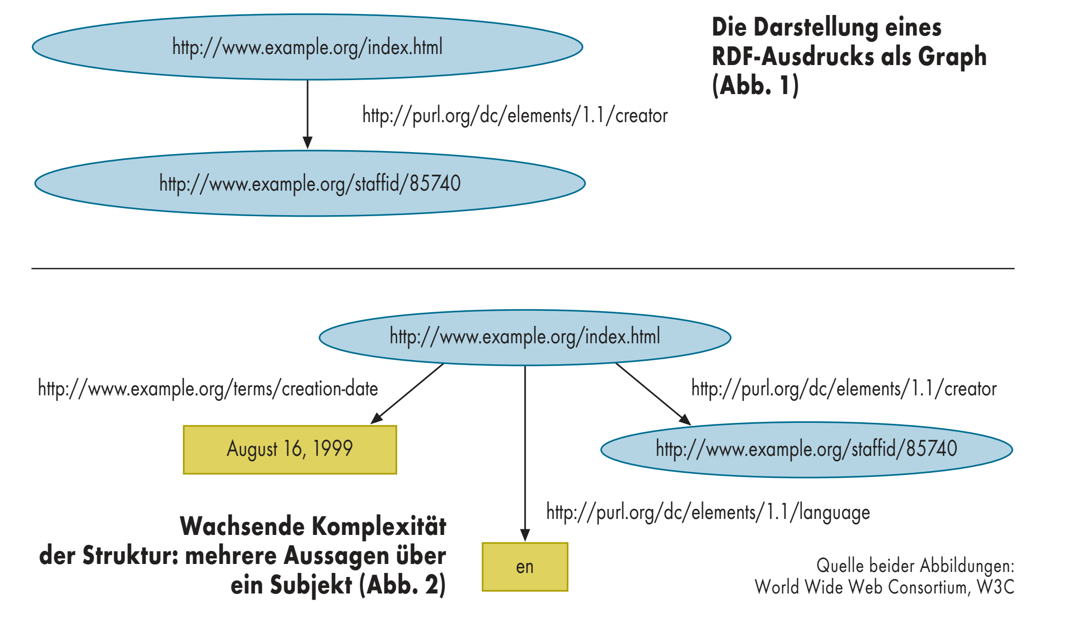
```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?urheber
WHERE
{
  <http://www.example.org/index.html>
  <http://purl.org/dc/elements/1.1/creator> ?urheber .
}
```

```
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.w3.org/2007/SPARQL/result.xsd">
  <head>
    <variable name="urheber"/>
  </head>
  <results>
    <result>
      <binding name="urheber">
        <uri>http://www.example.org/staffid/85740</uri>
      </binding>
    </result>
  </results>
</sparql>
```

Listing 2: Komplexere Anfrage/Ergebnis

```
PREFIX smwCat:  <http://ontoworld.org/wiki/Special:URIResolver/Category-3A>
PREFIX smwProp: <http://ontoworld.org/wiki/Special:URIResolver/Property-3A>
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?city ?country ?population ?geoLocation
WHERE {
  ?city rdf:type smwCat:City.
  ?city smwProp:Located_in ?country.
  ?city smwProp:Population ?population.
  ?city smwProp:Coordinates ?geoLocation.
}
```

```
<sparql
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xs="http://www.w3.org/2001/XMLSchema#"
    xmlns="http://www.w3.org/2005/sparql-results#" >
  <head>
    <variable name="city"/>
    <variable name="country"/>
    <variable name="population"/>
    <variable name="geoLocation"/>
  </head>
  <results>
    <result>
      <binding name="city">
        <uri>http://ontoworld.org/wiki/Special:URIResolver/Berlin</uri>
      </binding>
      <binding name="country">
        <uri>http://ontoworld.org/wiki/Special:URIResolver/Germany</uri>
      </binding>
      <binding name="population">
        <literal datatype="http://www.w3.org/2001/XMLSchema#double">3391407</literal>
      </binding>
      <binding name="geoLocation">
        <literal datatype="http://www.w3.org/2001/XMLSchema#string">52°31'0?N, 13°24'0?E</literal>
      </binding>
    </result>
    ...
  </results>
</sparql>>
```



Die Darstellung eines RDF-Ausdrucks als Graph (Abb. 1)

Wachsende Komplexität der Struktur: mehrere Aussagen über ein Subjekt (Abb. 2)

Quelle beider Abbildungen: World Wide Web Consortium, W3C

Listing 3: UNION

```
SELECT ?subject ?ingoingPredicate ?outgoingPredicate ?object
WHERE {
 {
  ?subject ?ingoingPredicate
  <http://ontoworld.org/wiki/Special:URIResolver/Berlin>.
 } UNION {
  <http://ontoworld.org/wiki/Special:URIResolver/Berlin>
  ?outgoingPredicate ?object.
 }
}
```

```
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
  <head>
  </head>
  <boolean>true</boolean>
</sparql>
```

Listing 4: ASK

```
PREFIX smwCat:  <http://ontoworld.org/wiki/Special:URIResolver/Category-3A>
PREFIX smwProp: <http://ontoworld.org/wiki/Special:URIResolver/Property-3A>
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
ASK {
  ?city rdf:type smwCat:City.
  ?city smwProp:Population ?population.
  FILTER ( ?population > 1000000 )
}
```

Listing 5: CONSTRUCT

```
PREFIX smwCat:  <http://ontoworld.org/wiki/Special:URIResolver/Category-3A>
PREFIX smwProp: <http://ontoworld.org/wiki/Special:URIResolver/Property-3A>
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex:      <http://example.org/meineOntologie/>
CONSTRUCT {
  ?city rdf:type ex:Stadt.
  ?city ex:istTeilVon ?country.
  ?city ex:hatEinwohnerZahl ?population.
}
WHERE {
  ?city rdf:type smwCat:City.
  ?city smwProp:Located_in ?country.
  ?city smwProp:Population ?population.
}
```

Listing 6: DESCRIBE

```
PREFIX smwCat:  <http://ontoworld.org/wiki/Special:URIResolver/Category-3A>
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
DESCRIBE ?x
WHERE {
  ?x rdf:type smwCat:City.
}
```

Listing 7: Joseki-Konfiguration

```
@prefix swivt:  <http://semantic-mediawiki.org/swivt/1.0#> .
@prefix smwCat: <http://ontoworld.org/wiki/Special:URIResolver/Category-3A> .
@prefix wiki:   <http://ontoworld.org/wiki/Special:URIResolver/> .
@prefix obowl:  <http://www.xspan.org/obo.owl#> .
@prefix rdfs:   <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf:   <http://xmlns.com/foaf/0.1/> .
@prefix :       <http://www.xspan.org/obo.owl#> .
@prefix owl:    <http://www.w3.org/2002/07/owl#> .
wiki:Halle_-28Saale-29
      a       swivt:Subject , smwCat:City ;
      rdfs:isDefinedBy
           <http://ontoworld.org/wiki/Special:ExportRDF/Halle_(Saale)> ;
      rdfs:label "Halle (Saale)" ;
      wiki:Property-3AAdjacent_to
              wiki:Saale ;
      wiki:Property-3AArea-23m-C2-B2
              "135000000"^^<http://www.w3.org/2001/XMLSchema#double> ;
      wiki:Property-3ALocated_in
              wiki:Germany , wiki:Saxony-2DAnhalt ;
      wiki:Property-3APopulation
              "240000"^^<http://www.w3.org/2001/XMLSchema#double> ;
      swivt:page <http://ontoworld.org/wiki/Halle_(Saale)> .
...
```

Listing 1. Offenkundig lassen sich auf diese Weise leicht komplexere Queries bilden, siehe Listing 2 oben.

Diese Anfrage zeigt, dass SPARQL Präfixe für häufig genutzte URIs definieren kann, was für eine deutlich bessere Lesbarkeit sorgt. Das Ergebnis der Anfrage ist auf Basis der Ontoworld-Daten die Menge aller dort erfassten Städte (*rdf:type smwCat:City*), des Landes (*smwProp:Located_in ?country*), in dem die jeweilige Stadt liegt, der Bevölkerungszahl (*?population*) und der Geokoordinaten (*?geoLocation*) – siehe Listing 2 unten.

Außer URIs kommen in der Antwort Literale vor. RDF erlaubt literale Knoten in einem Graphen, die – wie in diesem Beispielcode – mit einem Datentyp versehen sein können.

UNION fasst mehrere Queries in einer Anfrage zusammen, beispielsweise wie in Listing 3. Hierauf liefert eine SPARQL-Engine alle im Datenbestand enthaltenen RDF-Tripel, in denen *http://ontoworld.org/wiki/Special:URIResolver/Berlin* entweder als Subjekt oder als Prädikat vorkommt.

Prüfung auf Pattern

Um zu prüfen, ob bestimmte Daten vorhanden sind, ist das *ASK*-Statement vorgesehen. Im Gegensatz zu *SELECT* prüfen *ASK*-Anfragen lediglich, ob das angegebene Graphen-Pattern vorhanden ist, und sie liefern einen entsprechenden Boolschen Wert zurück (siehe Listing 4).

Die Frage lautet in diesem Fall: Gibt es im Datenbestand eine Instanz vom Typ *City* mit einer Einwohnerzahl von mindestens einer Million? Die Antwort findet sich in der unteren Hälfte von Listing 4. Eine Filterfunktion wie in der *ASK*-Anfrage erlaubt die Einschränkung der Ergebnismenge. Abgesehen von solchen logischen Vergleichen stellt *FILTER* weitere Optionen zur Verfügung, beispielsweise die Verwendung regulärer Ausdrücke. Weitere Informationen dazu finden sich in der W3C-Empfehlung.

Neben diesen reinen Anfragetechniken erlaubt es SPARQL, RDF-Graphen zurückzugeben beziehungsweise beliebige neue zu konstruieren. Es ist keine Überraschung, dass zu Letzterem *CONSTRUCT* dient. Listing 5 konstruiert aus den Informationen aller Städte einen neuen RDF-Graphen. Dieser wird allerdings mit einem anderen Vokabular umgesetzt, in diesem Sinne mit der (fiktiven) Ontologie hinter *http://example.org/meineOntologie/*.

Beschreibung über Ressourcen

Wenn nicht die Notwendigkeit besteht, die erfragten Daten in ein neues Vokabular zu überführen, kann man mit *DESCRIBE* arbeiten. So gibt

```
DESCRIBE <http://ontoworld.org/
wiki/Special:URIResolver/Berlin>
```

einen Graphen mit allen Daten des Wikis über Berlin zurück. Es lassen sich mit DESCRIBE aber auch komplexere Anfragen formulieren (siehe Listing 6). Die Empfehlung des W3C gibt nicht vor, in welchem Format die Antwort geliefert wird. Im Gegensatz zu anderen SPARQL-Queries, bei denen der Client den Aufbau des RDF kennen muss, um eine vernünftige Anfrage zu formulieren, dient *DESCRIBE* dazu, eine Beschreibung über eine Ressource im RDF-Vokabular der Datenquelle zu bekommen. Übersetzt heißt es so viel wie „Beschreib mal die Ressource X mit deinen Worten“. *DESCRIBE* kann damit ein guter erster Schritt zur Wissensentdeckung in beliebigen RDF-Quellen sein. Stellt man die obige Anfrage mit Joseki [f], bekommt man die Antwort im N3-Format (siehe Listing 7). Einer *WHERE*-Klausel

Onlinequellen

RDF und SPARQL beim W3C

[a] RDF Primer	www.w3.org/TR/rdf-primer/
[b] SPARQL Query Language for RDF	www.w3.org/TR/rdf-sparql-query/
[c] SPARQL Query Results XML Format	www.w3.org/TR/rdf-sparql-XMLres/
[d] SPARQL Protocol for RDF	www.w3.org/TR/rdf-sparql-protocol/

SPARQL-Software

Java:	
[e] ARQ	jena.sourceforge.net/ARQ/
[f] Joseki	www.joseki.org/
C:	
[g] Rasqal	librdf.org/rasqal/
PHP:	
[h] ARC	arc.semsol.org/
Common Lisp:	
[i] Twinql	www.holygoat.co.uk/projects/twinql/
C#:	
[j] SemWeb	razor.occams.info/code/semweb/
Übersicht:	
[k] SparqlImplementations	esw.w3.org/topic/SparqlImplementations

Weitere Links

[l] Jena Toolkit	jena.sourceforge.net
[m] Ontoworlds Semantic Media Wiki	ontoworld.org/

können die Ausdrücke *ORDER BY*, *LIMIT*, und *OFFSET* folgen. Sie dienen dazu, die Ergebnismenge zu sortieren (*ORDER BY*), ihren Umfang zu begrenzen (*LIMIT*) und anzugeben, bei welcher Position der Ergebnismenge die Ausgabe beginnen soll (*OFFSET*). Damit lässt sich ein Blättern in der Ergebnismenge realisieren, in dem der *OFFSET* zunächst bei 1 beginnt und beispielsweise ein *LIMIT* von 10 verwendet. Bei der nächsten Query setzt man den *OFFSET* auf 11 und erhält die Treffer 11 bis 20 und so weiter.

Fazit

Schon diese ersten Codebeispiele zeigen: SPARQL ist eine leistungsfähige und relativ einfach zu erlernende Anfragesprache für RDF-Graphen. Für die Entwicklung von Semantic-Web-Anwendungen ist sie eine unverzichtbare Komponente. Erfreulicherweise liegen mittlerweile relativ ausgereifte Implementierungen, auch als Open Source, vor. (hb)

STEFAN MINTERT UND BASTIAN SPANNEBERG

sind Informatiker und entwickeln Semantic-Web-2.0-Software bei Linkwerk.com.

Literatur

[1] Stefan Mintert, Bastian Spanneberg; Semantic Web; Verstehen, das; Semantische Erweiterung für MediaWiki; *iX* 11/2007, S. 102

[2] Michael Hausenblas; Semantic Web; Eulen nach Jena; Auf RDF-Basis: das Jena-Toolkit; *iX* 11/2005, S. 147

Mit SPARQL arbeiten

Die Codebeispiele des Artikels verwenden die exportierten RDF-Daten aus dem Ontoworld-Wiki. Wer auf dem eigenen Rechner mit SPARQL experimentieren möchte, kann beispielsweise das Jena Semantic Web Framework einsetzen [l] und die Daten dort speichern. Jena ist ein Open-Source-Java-Framework von Hewlett-Packard zur Entwicklung von Semantic-Web-Anwendungen. Die folgende Beschreibung soll helfen, die ersten Schritte mit Jena ohne Stolpern zu meistern – siehe außerdem den *iX*-Artikel [2].

Abgesehen von Jenas mächtiger API, die hier kein Thema ist, lässt sich die Software als Kommandozeilen-Werkzeug benutzen, um eine ganze Reihe von Datenbanken (MySQL, PostgreSQL, Oracle, Derby, …) als Persistenzschicht für RDF-Daten zu verwenden. Die weiteren Ausführungen gehen davon aus, dass eine lauffähige MySQL-Installation vorhanden ist.

Standard-DB-Schema noch zu bevorzugen

Über das derzeit bei Jena enthaltene Datenbank-Schema hinaus arbeiten die Entwickler gerade an SDB (*jena.sourceforge.net/SDB*), einer neuen Triple-Store-API inklusive einem neuen Datenbank-Schema, das insbesondere für komplexere SPARQL-Anfragen optimiert ist. SDB befindet sich noch in einem frühen Entwicklungsstadium. Deshalb ist im Moment das Standardschema zu empfehlen. Joseki allerdings kann schon jetzt mit SDB-Datenbeständen arbeiten.

Der erste praktische Schritt besteht in der Erzeugung einer passenden Datenbank. Nach dem Anlegen einer leeren Datenbank in MySQL initialisiert der folgende Aufruf im Jena-Verzeichnis die Datenbank mit den richtigen Einstellungen. Die Aufrufsyntax geht von einem Windows-System aus. Unter Unix ist die Syntax für den *Classpath* geeignet zu wählen (*-cp .:lib/**).

```
java -cp .;lib\* jena.dbcreate \
 -db jdbc:mysql://localhost/jena \
 -dbType MySQL  -dbUser dbUser \
 -dbPassword dbPassword -model ix
```

Über den Parameter *model* lässt sich der Name der zu erzeugenden Datenbank, im Jena-Jargon „Model“ genannt, angeben. Anschließend lädt die nächste Anweisung die benötigten RDF-Daten in dieses Modell:

```
java -cp .;lib\* jena.dbload \
 -db jdbc:mysql://localhost/jena \
 -dbType MySQL -dbUser root \
 -dbPassword mysqlroot \
 -model ix c:\sparql\ontoworld.xml
```

Jena kann nun SPARQL-Anfragen auf dem Datenbestand ausführen. Dazu stehen zwei Wege offen: Die zu Jena gehörende SPARQL-Engine ARQ [e] kann ein selbst geschriebenes Java-Programm ansprechen, oder man nutzt Joseki [f], eine weitere, separat entwickelte Komponente aus dem Jena-Universum.

Joseki ist ein SPARQL-Query-Server und läuft als Webservice in einem integrierten Jetty. Bei Bedarf lässt sich Joseki außerdem als Webanwendung in anderen Servlet-Containern benutzen. Der Service kann sowohl mit RDF-Daten in Dateien als auch in Datenbanken umgehen. Er implementiert das SPARQL-Protokoll und liefert Ergebnisse gemäß der SPARQL Query Results Proposed Recommendation. Beim Entpacken von Joseki ist darauf zu achten, dass im Pfad keine Leerzeichen enthalten sind, da der Server anderenfalls nicht ordnungsgemäß startet.

Josekis Konfiguration erfolgt ebenfalls in RDF. Da die Konfigurationsdatei *joseki-config.ttl* für nicht RDF-affine Nutzer etwas verwirrend sein kann, hier ein kurzer Abriss der für die vorliegenden Codebeispiele wesentlichen Konfiguration (siehe Listing 8).

Einrichten von Diensten mit Joseki

Unter Joseki kann man mehrere sogenannte Services einrichten. Die wesentlichen Angaben für den jeweiligen Service finden sich in den Prädikaten *serviceRef*, *dataset* und *processor*. Die Angabe in *serviceRef* dient dazu, ankommende Requests auf die passenden Datenbestände abzubilden.

Beim Anlegen von eigenen Services ist daher ein entsprechender Eintrag in Josekis *web.xml* zu tätigen. Um den Einstieg zu erleichtern, bleibt das Listing beim Namen *books* aus der Default-Konfiguration. Das ermöglicht die Nutzung des mitgelieferten Query-Formulars, in dem dieser Name fest eingebrannt ist. Das Prädikat *dataset* verweist auf eine weiter unten folgende Definition. Dort gibt es eine Referenz auf den zu diesem Dataset gehörenden Datenbestand. Die Angabe unter *processor* verweist schließlich auf die Implementierung des SPARQL-Query-Prozessors. Es ist konfigurierbar, ob der abzufragende Graph in einer Query angegeben werden kann, oder ob, wie im vorliegenden Fall, Abfragen nur gegen einen vorgegebenen Datenbestand möglich sind. Sind alle diese Angaben korrekt gemacht, lässt sich der Server von der Kommandozeile starten:

```
set JOSEKIROOT=C:\Joseki-3.1
bin\joseki_path
bin\rdfserver
```

Der letzte Aufruf muss im Joseki-Verzeichnis stattfinden, da das Tool sonst die Konfigurationdatei *joseki-config.ttl* nicht finden kann. Jetzt steht das integrierte Query-Formular unter *http://localhost:2020/query.html* zum Testen von Queries zur Verfügung. Die Nutzung des Formulars bietet den Vorteil, dass auf das zurückgegebene XML gleich ein XSLT-Stylesheet angewandt werden kann, das die Antwort in einer HTML-Seite darstellt.

Listing 8: joseki-config.ttl

```
## This file is written in N3 / Turtle
@prefix rdfs:   <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd:    <http://www.w3.org/2001/XMLSchema#> .
@prefix module: <http://joseki.org/2003/06/module#> .
@prefix joseki: <http://joseki.org/2005/06/configuration#> .
@prefix ja:     <http://jena.hpl.hp.com/2005/11/Assembler#> .
## About this configuration
<> rdfs:label "Joseki Configuration File" .
## About this server
[]  rdf:type joseki:Server ;
    joseki:initialization
        [ module:implementation
            [ module:className <java:org.joseki.util.ServiceInitSimple> ;
              rdfs:label "Example initializer" ; ]
        ] ;
    .
## Service 1 - SPARQL processor for the ontoworld dataset
[]  rdf:type            joseki:Service ;
    rdfs:label          "SPARQL service for the ontoworld model" ;
    joseki:serviceRef   "books" ;
    joseki:dataset      _:ix ;
    joseki:processor    joseki:ProcessorSPARQL_FixedDS ;
    .
## Datasets
_:ix   rdf:type         ja:RDFDataset ;
       ja:defaultGraph  _:ixModel ;
       rdfs:label       "DBMS iX Dataset" ;
    .
## Models
_:ixModel rdf:type ja:RDBModel ;
    rdfs:label "iX SPARQL" ;
    ja:connection
    [
        ja:dbType         "MySQL" ;
        ja:dbURL          "jdbc:mysql://localhost:3306/jena" ;
        ja:dbUser         "dbUser" ;
        ja:dbPassword     "dbPassword" ;
        ja:dbClass        "com.mysql.jdbc.Driver" ;
    ] ;
    ja:modelName "ix"
    .
## Processors
joseki:ProcessorSPARQL_FixedDS
    rdfs:label "SPARQL processor for fixed datasets" ;
    rdf:type joseki:Processor ;
    module:implementation joseki:ImplSPARQL ;
    # This processor does not accept queries with FROM/FROM NAMED
    joseki:allowExplicitDataset      "false"^^xsd:boolean ;
    joseki:allowWebLoading           "false"^^xsd:boolean ;
    joseki:lockingPolicy             joseki:lockingPolicyMRSW ;
    .
joseki:ImplSPARQL
    rdf:type   joseki:ServiceImpl ;
    module:className <java:org.joseki.processors.SPARQL> .
```

Continuations als Sprachmittel

Fortsetzungs-geschichte

Frank Müller



Die Entwicklung moderner Webapplikationen stellt den Programmierer vor diverse Herausforderungen. Eine davon ist die Überwindung der Zustandslosigkeit des HTTP. Ein aktuell stark diskutierter Lösungsansatz hierfür ist die Verwendung eines Framework auf Basis von Continuations.

Continuations sind einer der derzeit vieldiskutierten Ansätze, um die Zustandslosigkeit von HTTP in der Programmierung zu umschiffen. Dass dieses Sprachmittel aus einer Zeit weit vor dem Web stammt, ist vielen Beteiligten unbekannt. Dabei könnte die Kenntnis der Geschichte dieses Ansatzes den Blick für dessen Stärken und Schwächen schärfen.

Eingeführt wurde der Begriff Continuation bereits in den 60er-Jahren von Christopher Strachey, Christoper P. Wadsworth und John Reynolds. Das erste Web-Framework, das Continuations für den Kontrollfluss nutzte, war Seaside. Diese in Smalltalk entwickelte Bibliothek [1] verzichtet auf ein aufwendiges, in XML oder in anderer Form definiertes Mapping von URLs zu Komponenten. Stattdessen erzeugen Seaside-Komponenten ihre individuelle Ausgabe programmatisch durch die Methode *renderContentOn:*. Sie erhält als Argument ein HTML-Render-Objekt, das über Methoden zur Erzeugung der HTML-Ausgabe verfügt. Die Methoden für die Erzeugung von Links erhalten als Callback unter anderem einen Block, auch als Closure bekannt, für den nach einer Auswahl auszuführenden Code.

Zur Darstellung dieses Links generiert Seaside eine individuelle ID, die gemeinsam mit der Session-ID Teil der URL ist. Gleichzeitig wird der übergebene Callback-Block in einem Dictionary gesichert und steht dadurch bei einer erneuten Auswahl des Links wieder zur Verfügung. So kann der Programmfluss in einfachster Form fortgesetzt werden, sowohl in der aktuellen Komponente als auch beim Start einer neuen.

Blöcke lassen sich zwar wunderbar für eine spätere Ausführung speichern. Doch die erfolgt immer in einem Kontext, der sowohl alle lokalen Variablen im Sichtbereich des Blocks als auch die Aufrufkette, die zur Aktivierung der Methode führte, umfasst. Mit dem Verlassen der Methode, in der der Block definiert ist, verändert sich dessen Kontext. Das Ergebnis einer späteren Blockausführung auf Basis einer Benutzerin-



teraktion ist dann kaum vorhersehbar.

An eben dieser Stelle setzen Continuations an. Wie der Name bereits andeutet, handelt es sich um eine Technik zur Fortsetzung eines Programms an einer definierten Position. Dabei geht es nicht nur um einen einfachen Sprungbefehl, sondern um ein Konstrukt zur Sicherung des aktuellen Kontexts, zu dessen Übermittlung für eine Weiterverarbeitung an einen Block und zur Terminierung des aktuellen Kontexts mit dem Zweck, das Programm mit dem gesicherten Exemplar fortzusetzen. So ist das nutzende Programm in der Lage, beliebige Punkte in seiner Ausführung als Schnappschuss zu sichern und mittels einer virtuellen Zeitreise an diese Stellen zurückzuspringen.

Mehr als ein Sprung

Die Nutzung von Continuations ist aus programmtechnischer Sicht einfach. Die für die Erzeugung notwendige Methode erwartet wie bereits erwähnt einen Block mit einem einzigen Argument. Das enthält die aktuelle Continuation und lässt sich dann beispielsweise für die weitere Nutzung speichern. Im Block sind weitere Programmschritte möglich und das Ergebnis wird an den Aufrufer übergeben.

```
{{{
Transcript clear.
continuation := nil.
tmp := Continuation currentDo:
  [:cc | continuation := cc. false].
tmp ifFalse: [
    Transcript show:
        ,A => , , tmp asString; cr.
    continuation value: true.
    Transcript show:
        ,B => , , tmp asString; cr].
Transcript show:
    ,C => , , tmp asString; cr.
}}}
```

Die Ausgabe dieses Beispiels verdeutlicht mit

```
{{{
A => false
C => true
}}}
```

den Ablauf. Der Aufruf von *Continuation currentDo:* speichert den aktuellen Kontext in der Variable *continuation* und gibt false an *tmp* zurück. Das steuert den weiteren Ablauf, also die erste Ausgabe (A) und

den Aufruf der Continuation mit dem Wert *true*. Diese so ausgelöste Fortsetzung am Ort der Definition sorgt dafür, dass anschließend erstens die zweite Ausgabe (B) nicht mehr ausgeführt wird und zweitens *tmp* den Wert *true* erhält. Damit führt das Programm den Block nicht erneut aus und erreicht die dritte Ausgabe (C). Ohne die *if*-Abfrage würde dieses Beispiel mit seinen stetigen Rücksprüngen in einer Endlosschleife enden.

So einfach dieses abstrakte Beispiel ist, macht es doch bereits deutlich, mit welcher Vorsicht diese Fähigkeit zu nutzen ist. In gewissem Sinne ähneln Continuations einer funktionalen Form des Gotos, das nicht umsonst seit Edsger Dijkstras Aussage im Jahre 1968 als schädlich gilt [2]. Der Entwickler erhält die Möglichkeit, den Programmfluss abseits der regulären Anweisungsfolge zu verändern. Der entstehende Code ist schwerer zu debuggen und für neue Entwickler schlecht durchschaubar – beides keine unerheblichen Faktoren. Hilfsmittel zur Code-Inspektion, wie die Verfolgung von Aufrufern oder Implementierern von Methoden, greifen bei Continuations nicht. Dieser Zwiespalt zeigt sich auch im sparsamen Gebrauch dieses Features in den Bibliotheken. So basiert in Smalltalk bisher nur Seaside als bekanntes Framework auf diesem Verfahren.

Ohne Return: Scheme

Eine weitere Sprache, die nicht nur für ihre Continuations bekannt ist, sondern auch gerne für die Vermittlung dieses Konstrukts herangezogen wird, ist Scheme. So kennt diese funktionale Sprache kein Return-Statement, was die Entwicklung von Funktionen wie „Finde das erste passende Element in einer Liste.“ erschwert oder zumindest ineffizient macht. Mittels der Funktion *(call-with-current-continuation)* – oder seinem Alias *(call/cc)* – lässt sich jedoch leicht ein Ausweg aus diesem Dilemma implementieren. Wie in Smalltalk hat diese Funktion ein Argument. Dieses ist entweder wiederum eine Funktion mit einem Argument, wie bei Smalltalk zur Übergabe der aktuellen Continuation, oder die Continuation selbst. Im ersten Fall wird die Funktion aufgerufen, im zweiten die Continuation fortgesetzt. Damit lässt sich die eben genannte Funktion leicht implementieren:

```
{{{
(define (find fits? list)
  (call/cc
    (lambda (return)
      (for-each
        (lambda (element)
          (if (fits? element)
            (return element)))
        list)
      f)))
}}}
```

Der Aufruf der Continuation *return* mit dem gefundenen Element sorgt für eine Fortsetzung am Definitionspunkt, also außerhalb der Schleife und damit für die Rückgabe des Elements als Ergebnis. Passt hingegen kein Element der Liste, wird am Ende *false* zurückgegeben. Im Unterschied zum Smalltalk-Beispiel zeigt dieses Beispiel noch keinen Rücksprung an den Punkt der Continuation, sondern nur das Herausspringen aus der aktuellen Funktion. Sprünge in beide Richtungen erlauben weitere praktische Konstrukte. So ist es hiermit möglich, Funktionen mit einem Zustand auszustatten.

```
{{{
(define (make-counter start increment)
  ;; Schleife fuer Inkrement
  (define (increase)
    (let loop ((counter start))
      (call/cc (lambda (cc)
                 (set! increase
                   (lambda () (cc ,dummy)))
                 (caller counter)))
      (loop (+ counter increment))))
 ;; Funktion fuer jeweils einmaligen
 ;; Aufruf des Inkrements
  (define (caller)
    (call/cc (lambda (cc)
               (set! caller cc)
               (increase))))
  caller)
}}}
```

Dieses Beispiel sorgt für die Erzeugung eines Zählers mit einem Startwert sowie einem Inkrement. Die beiden Unterfunktionen rufen sich gegenseitig auf, definieren sich zur Realisierung eines Start-Stop-Verhaltens mit jedem Durchlauf um und sorgen dafür, dass die Schleife in *(increase)* einmal pro Aufruf durchlaufen wird. Damit erzeugt der Aufruf von *(define counter-five (make-counter 0 5))* die Funktion *(counter-five)*, die mit ihren Aufrufen die Zahlenfolge 0, 5, 10, 15, 20 et cetera ausgibt. Die Erzeugung einer zweiten Funktion *(define counter-two (make-counter 1 2))* liefert hingegen *(counter-two)* und damit die Zahlenfolge 1, 3, 5, 7, 9 ….. Auf diese Weise lassen sich beispielsweise IDs für Datenbankeinträge generieren.

Hin und her und zurück

Ein weiteres Beispiel für den Einsatz von Continuations ist der Sprung zwischen zwei Funktionen, auch als Koroutinen bekannt. Auf diese Weise lässt sich eine Form von paralleler Ausführung simulieren, ein kooperatives Multitasking. Hierfür werden zwei Funktionen mit je einem Argument, das für die Übernahme der jeweils anderen Funktion zuständig ist, definiert. Im Rahmen der internen Verarbeitung oder Schleifen rufen sich die Funktionen gegenseitig regelmäßig mit *(set! other-func (call/cc other-func))* auf. Dabei wird einerseits der aktuelle Stand gesichert und andererseits gleichzeitig der jeweils anderen Funktion für den Rückruf übergeben. Auf diese Weise lässt sich beispielsweise im Rahmen einer länger andauernden Berechnung oder Verarbeitung eine Information an den wartenden Benutzer ausgeben. Im folgenden Beispiel ist es ein kreiselnder Strich in Klammern, angezeigt als Uhr-Ersatz hinter jeder dargestellten Nummer.

```
{{{
(define (animated-loop num)
  (define (loop-func animator-func)
    (let loop ((counter 1))
      (display counter)
      (set! animator-func
          (call/cc animator-func))
      (if (< counter num)
          (loop (+ counter 1)))))
  (define (animator-func loop-func)
    (let loop ()
      (for-each (lambda (animation)
            (display animation)
            (newline)
            (set! loop-func
                 (call/cc loop-func)))
                '(" (|)" " (/)" " (-)" " (\\)"))
      (loop)))
  (loop-func animator-func))
}}}
```

Diese Vereinfachung fasst die beiden Funktionen für die Schleife und ihre Animation zusammen und kann mit *(animated-loop 10)* die Werte von 1 bis 10 mit einer Uhr anzeigen. Alternativ kann es der Entwickler der langlaufenden Funktion dem Nutzer freistellen, wie er seine Benutzerinformation gestalten möchte.

Sprachen und Anwendungen

Außer Smalltalk und Scheme verfügen weitere Sprachen über Continuations. Zu den bekannteren gehört beispielsweise Ruby, das mit *callcc { |cc| … }* über die gleiche Mächtigkeit verfügt. Dazu kommen noch Haskell, Factor, Parrot, ML, Rhino, Scala, Pico und auch C. Prinzipiell lässt sich ein als Continuation Passing Style bekanntes Verfahren in jeder Sprache implementieren, die über Funktionsobjekte verfügt. In diesem Fall wird nicht das Ergebnis einer Funktion zurückgegeben und weiter verwendet, sondern der Funktion ein Funktionsobjekt übergeben, das nach erfolgter Berechnung das Ergebnis erhält. Damit lassen sich viele der Vorteile vollwertiger Continuations ebenfalls realisieren. Ein weiterer Weg besteht in Java-Bibliotheken zur Manipulation des Bytecodes.

Die Einleitung mit Smalltalk deutete bereits ein bevorzugtes Anwendungsfeld für Continuations an. Es sind die Web-Frameworks, losgetreten durch den Erfolg von Seaside [b]. Sicherlich ist die Basis der auf Seaside basierenden Anwendungen im Verhältnis zum Mainstream mit Java und .Net gering, als Impulsgeber für neue Lösungsansätze hat es hingegen Bedeutung und Berühmtheit erlangt. Allerdings zeigt Seaside auch einen Nachteil der Continuations in Szenarien wie Webanwendungen. Die Sicherung des jeweils aktuellen Kontexts ist speicherintensiv und belastet ebenso wie die Restauration die VM. Skalierbare Applikationen erfordern daher den Einsatz paralleler Images sowie eines Load Balancings.

Bei den Webbibliotheken steht Seaside inzwischen nicht mehr alleine da. So hat auch !PLT Scheme [c] Continuations in seine Scheme Web Servlets Library integriert, wenn auch nicht in der Konsequenz und mit dem Reichtum an vorgefertigten Komponenten wie Seaside. Ein weiteres Scheme Web Framework basierend auf den Ideen von Seaside ist der Modal Web Server für Chicken Scheme. Die Entwicklung dieser Bibiothek ist jedoch eher experimentell und wird nicht weiter verfolgt. Für den großen Scheme-Bruder Common Lisp bietet das Framework UnCommon Web [d] die entsprechenden Funktionen.

Ruby auch ohne Rails

Trotz des Erfolg von Ruby on Rails gibt es auch für diese Sprache Projekte, die Web-Frameworks auf Basis von Continuations realisieren. Eines ist Borges [e], das sich stark an Seaside orientiert. Daneben existiert noch Wee, das allerdings eingeschlafen ist. Aktiver ist hingegen Perls Jifty [f]. In Python startete Subway einen Versuch, sich auf Basis von Continuations als Alternative durchzusetzen. Leider ist dieses Projekt ebenfalls sanft entschlummert.

Auch in der Java-Welt bedient man sich wie oben bereits angedeutet der Continuations für das Web. Das Framework der Wahl trägt den Namen RIFE [g] und stellt eine vollwertige Umgebung sowohl mit Templates als auch mit Continuations und vielen weiteren Features zur Verfügung. Wettbewerber sind seitens Spring das Unterprojekt Web Flow [h] sowie Apache Cocoon [i] mittels Flow Script.

Fazit

Die Nutzung von Continuations ermöglicht die Ergänzung von Sprachen um ein Mittel, das ihrem eigentlichen Paradigma nicht entspricht. So brechen sie sowohl aus der Objektorientierung als auch aus der funktionalen Programmierung aus. Genau darin liegen Reiz und Risiko. Die aufgeführten Beispiele geben einen Einblick, wie Continuations helfen, kleinere Aufgaben einfach zu lösen. Von der Wirkung ähnliche Ansätze ohne den Einsatz von Continuations wären aufwendiger. Auf der anderen Seite sind diese jedoch für die meisten Entwickler einfacher nachzuvollziehen. Dies bedeutet, dass der reguläre Code in der Regel für Teams mit wechselnder Besetzung – also im Alltag – wartbarer wird. Hierin mag auch einer der Gründe liegen, warum diverse auf Continuations beruhende Projekte wieder eingeschlafen sind.

Dennoch kann sich im anspruchsvollen Fall, wie in der Entwicklung der Web-Frameworks in Smalltalk, Scheme oder Python, der Einsatz lohnen. Continuations erlauben hier nicht nur die Erweiterung der Sprachmittel, sondern auch die Überwindung von Schwächen des Lösungsumfelds. Bei Webanwendungen ist dies die Zustandslosigkeit des Protokolls, die ohne Continuations durch eine Menge gespeicherter Session-Variablen mehr oder minder geschickt umgangen wird. Continuations hingegen ermöglichen es, echte Zustände zu sichern und zu restaurieren. Dies erlaubt einen natürlichen Programmfluss ohne Bruch.

Die Komplexität konzentriert sich hier auf einen kleinen Bereich im Framework, abgeschirmt vom Entwickler der Webanwendung und gepflegt durch den Spezialisten. Die Beispiele in Seaside zeigen dies. Aus Gründen der Wartbarkeit sollte der Einsatz dennoch sparsam und wohl kommentiert sein. (JS)

FRANK MÜLLER

arbeitet als Teamleiter und Senior Consultant bei der BTC Business Technology Consulting AG in Oldenburg.

Literatur

[1] Frank Müller; Smalltalk; Ufer gewinnen, Webanwendungen mit Squeak und Seaside; *iX* 5/07; S. 136

[2] Edsger Dijkstra; Go To Statement Considered Harmful; Communications of the ACM 11; S. 147

Onlinequellen

[a] david.tribble.com/text/goto.html

[b] www.seaside.st

[c] www.plt-scheme.org

[d] common-lisp.net/project/ucw

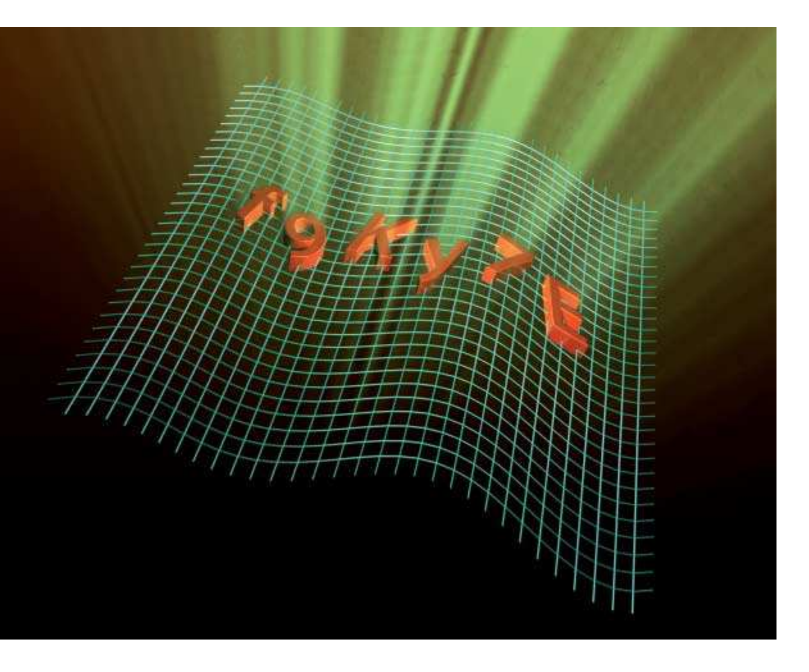
[e] borges.rubyforge.org

[f] jifty.org

[g] www.rifers.org

[h] www.springframework.org/webflow

[i] cocoon.apache.org



Captchas in PHP

Mensch oder Maschine

Hubert Benjamin Ritzdorf

Ein Turing-Test soll zwischen einem menschlichen und einem Elektronengehirn unterscheiden. Captchas versuchen damit Webangebote vor Bots zu schützen.

Fast jeder hat sie schon mal gesehen und sich gefragt, welchen Zweck sie eigentlich erfüllen: Kleine Bilder mit verschwommenen, verzerrten Zeichen vor beliebigem Hintergrund, die der Besucher der Webseite in ein Textfeld schreiben soll. Es sind sogenannte Captchas – „Completely Automated Public Turing test to tell Computers and Humans Apart“, also ein automatischer, öffentlicher Test zur Unterscheidung zwischen Mensch und Machine.

Das Prinzip klingt einfach und wirksam. Schließlich ist es für jeden Menschen ein Leichtes, den verzerrten Text aus dem Bild einzutippen, wohingegen der Computer diese Fähigkeit nicht besitzt. So kommen Captchas überall dort zum Einsatz, wo automatisierte Software Schaden anrichten könnte: in Foren und Gästebüchern zum Schutz vor automatischen Werbeeinträgen, bei Abstimmungen, bei der Beantragung neuer E-Mail-Adressen und an vielen weiteren Stellen.

Zu den positiven Aspekten dieser Technik gehört neben einem weitgehenden Schutz die einfache Implementierung, im Folgenden mit PHP veranschaulicht. Negativ ist anzumerken, dass ein Captcha mittlerweile kein Turing-Test im engen Sinne mehr ist, da Computer Text in Bildern mit wachsendem Erfolg erkennen können.

Bei der Implementierung von Captchas helfen verschiedene PHP-Erweiterungen, etwa die noch kaum dokumentierte Imagick-Image-Library, oder die besser beschriebene GD-Library, die jedoch weniger Funktionen bietet. Deshalb fiel die Wahl auf Imagick, das sich mit dem Shell-Kommando *pecl* nachinstallieren lässt. Es stellt die Klassen *Imagick, ImagickDraw, ImagickPixel* und *ImagickPixelIterator* bereit. Im folgenden Beispiel (siehe Listing 1) spielen nur die drei ersten eine Rolle.

Vorsicht bei der Schriftauswahl

Zu Beginn erzeugt es die nötigen Objekte, setzt den Hintergrund auf Weiß und stellt die Schriftart für das *ImagickDraw*-Objekt ein. Mehr Sicherheit bieten Captchas, wenn das Skript die Schrift für jedes Zeichen zufällig bestimmt. Allerdings ist darauf zu achten, dass sie in allen möglichen Varianten lesbar bleiben. Ein Wechsel zwischen mageren und fetten Schriften empfiehlt sich nicht, da die fetten Buchstaben mehr Platz benötigen und durch Überlagerung die Lesbarkeit leidet.

Dem Setzen der Schriftgröße folgt das Erzeugen des anzuzeigenden Zufallstexts. Von der zufälligen Auswahl aus Wörterbüchern ist dabei abzuraten, da dies Angreifern die Arbeit erleichtert. Stattdessen sollte man eine Menge von Großbuchstaben und Zahlen festlegen, die das Captcha enthalten darf. 1, I und J oder 0, Q und O eignen sich nicht, da auch normalsichtige Besucher sie schwer unterscheiden können. Mischt man beispielsweise den String aus allen erlaubten Zeichen und liest ab einer Zufallsposition sechs Zeichen aus, sollte das für eine solche Anwendung ausreichend sicher sein.

Nun erzeugt das *Imagick*-Objekt ein neues 80 × 30 Pixel großes Bild. Es hat einen weißen Hintergrund und darüber einen mit *addNoiseImage()* erzeugten Haufen bunter zufälliger Pixel, sogenanntes Bildrauschen. Anschließend schreibt *annotateImage()* den Text hinein. Es ist zwar möglich, den gesamten Zufallsstring auf einmal auszugeben, doch kann man unaufwendig jedem Buchstaben eine zufällige vertikale Position geben (3. Parameter der Funktion) und ihn drehen (4. Parameter). Aufrufe von *swirl()* und *wave()* modifizieren den Text durch Drehen um die Bildmitte und wellenförmiges Verzerren. Damit ist das Captcha fertig. Die *wave()*-Funktion hat es gegenüber der ursprünglichen Vorgabe etwas vergrößert. Wer eine feste Größe benötigt, kann das Bild mit *cropImage()* zurechtschneiden.

Ein letzter Bearbeitungsschritt zeichnet vier zufällige Linien durch das Bild. Es empfiehlt sich, die Endpunkte mindestens einer Linie einzugrenzen, da sonst alle am oberen Rand liegen könnten. An dieser Stelle ist zu bemerken, dass es sicherlich immer „bessere“ und „schlechtere“ Captchas gibt, also sehr leicht zu entschlüsselnde und andere, die sogar für einen Menschen ein Rätsel darstellen. Es gilt, diese Extreme zu vermeiden und andererseits die nötige Variationsvielfalt zu schaffen, die ein allzu leichtes Entschlüsseln verhindert. In diesem Fall kann der Entwickler zum Beispiel durch Beschränkung der Zufallswerte sicherstellen, dass immer eine Linie von links unten nach rechts oben verläuft.

Nach dem Zeichnen der Linien legt das Skript das Bildformat als JPEG fest. Vor der Übergabe an den Browser muss es den Text speichern, um später die Eingabe überprüfen zu können. Das Speichern übernimmt üblicherweise eine Session-Variable, die erst nach dem Starten einer Session verfügbar ist.

Mit *<img src=“captcha.php“>* bindet man das Captcha in ein HTML-Formular ein. Da möglicherweise auch ein Mensch den Text nicht erkennt, sollten Anwender ein neues Captcha generieren können. Des Weiteren benötigt man ein Feld für die Eingabe des Texts und einen Knopf zum Abschicken des Formulars.

Session speichert das Gesuchte

Zum Vergleich des eingegebenen mit dem gespeicherten Text ist zunächst die Session fortzusetzen, die den korrekten String enthält. Falls er noch existiert (die Session also noch gültig ist) und mit dem eingegebenen String übereinstimmt, sieht man den Captcha-Test als bestanden an. Nun sind die geschützten Aktionen zugänglich.

Nach einem bestandenen Test hat die Session ihren Dienst getan und muss beendet werden. Sonst könnte ein Mensch einmal das Captcha lösen und anschließend einem Bot die Session-ID samt Lösung überlassen, sodass der beispielsweise noch Tausende Male abstimmte könnte. Für das Beenden der Session reicht es hier, die Variable zu löschen, die den Code enthält. Das Skript bemerkt gegebenenfalls bei der Überprüfung am Anfang, dass die übergebene Session bereits beendet ist und lehnt die neue Anfrage ab.

Texterkennungssoftware, sogenannte OCR-Programme, vermag mittlerweile Text aus Bildern zu lesen. Auf das Knacken von Captchas spezialisierten Weiterentwicklungen dieser Programme gelingt die Entzifferung immer besser. Im Erkennen einzelner, verfremdeter Zeichen haben einige den Menschen bereits überholt. Auf Internetseiten wie Captcha Killer, AICaptcha, Breaking a Visual Captcha und PWNtcha kann man den Fortschritt dieser Verfahren verfolgen. Wie fast immer bei Sicherheitstechniken entwickelt sich ein Wettlauf: Während Entwickler bessere Captchas erstellen, gibt es stärkere Bemühungen, sie zu umgehen. Der Aufwand für die Seitenbetreiber steigt und so hat die Suche nach Alternativen begonnen. Matt May vom W3-Consortium hat einige mit ihren Stärken und Schwächen zusammengestellt. So wären etwa Rechenaufgaben oder Fragen verwendbar: Was ist ein Planet? a) Baum b) Erde c) Hund. Doch auch dies ist nicht die Ultima Ratio, da nur extrem viele, stets aktualisierte Fragen Maschinen langfristig fernhalten können. Auch Sound-Captchas scheinen wenig Erfolg versprechend, bräuchte man doch Lautsprecher an jedem PC und eine gute Tonqualität.

Sinnvoll wäre, das Bild eines bekannten Tieres, etwa eines Hundes, zu verwenden und nach der Tierart zu fragen. Viele verschiedene Bilder pro Tierart erschweren die Arbeit für Maschinen, stellen für Menschen aber keine Hürde dar. Andererseits scheitern viele fremdsprachige Besucher an der Aufgabe, denen die Tierarten in einer anderen Sprache nicht bekannt sind.

Viele Lösungen weisen noch Schwächen in Bezug auf Barrierefreiheit auf. So bleibt etwa Blinden all das versperrt, was durch Captchas geschützt ist. Das hat in den USA bereits erste Anwälte auf den Plan gerufen. Dieser Probleme hat sich das Captcha-Projekt angenommen. Es bietet Plug-ins, mit denen sich grafische oder akustische Captchas auf der eigenen Website einsetzen lassen. Sie dienen einem guten Zweck: Als grafische Captchas fungieren bei der Digitalisierung von Büchern nicht erkannte Stellen. Dabei werden bereits entschlüsselte Ausdrücke, die zur Verifizierung der Eingabe dienen, mit noch unbekannten kombiniert. Jede Entzifferung stellt auf diese Weise einen Beitrag zum Erhalt alter Bücher dar.

Fazit

Als einfacher Schutzmechanismus für durchschnittlich besuchte Seiten ist ein Captcha gut geeignet. Leicht zu implementieren, bieten sich dank der verschiedenen Parameter viele Möglichkeiten der individuellen Gestaltung. Zudem muss der Angreifer den erheblichen Mehraufwand eines OCR-Systems auf sich nehmen, was viele abschrecken dürfte. Die Entwicklungen beider Gruppen laufen auf Hochtouren. So wird es bald ein PHP-Objekt für Captchas geben, das die notwendigen Methoden enthält und eine noch leichtere Einbindung bieten soll. Allerdings wurden auch hier bereits Sicherheitslücken entdeckt. (ck)

Listing 1: Captcha mit PHP erzeugen

```
<?php
/* Imagick-Objekt erzeugen */
$Imagick = new Imagick();
/* Weißer Hintergrund */
$bg = new ImagickPixel();
$bg->setColor('white');
/* ImagickDraw-Objekt erzeugen */
$ImagickDraw = new ImagickDraw();
/* Schrift setzen */
$ImagickDraw->setFont('/path/to/fonts/font.ttf');
$ImagickDraw->setFontSize(20);
/* Zufallstext erzeugen */
$zeichen = 'ACBXZRMYHTL23456789';
$randomnr = mt_rand(0, strlen($zeichen)-6);
$captcha_text = substr(str_shuffle($zeichen),
   $randomnr, 6);
/* Neues Bild mit weißem Hintergrund erzeugen */
$Imagick->newImage(85, 30, $bg);
/* Bildrauschen erzeugen */
$Imagick->addNoiseImage(imagick::NOISE_IMPULSE);
/* Text ins Bild schreiben */
for($i=0; $i<6; $i++){
   $randomnr = mt_rand(15,25);
   $Imagick->annotateImage($ImagickDraw, 12*$i,
      $randomnr, ($randomnr-20)*2, $captcha_text{$i});
}
/* Bild verzerren */
$Imagick->swirlImage(30);
$Imagick->waveImage(4, 60);
/* Zufällige Linien erzeugen */
$ImagickDraw->line(mt_rand(0, 20), mt_rand(20, 30),
    mt_rand(50, 70), mt_rand(0, 10));
$ImagickDraw->line(mt_rand(0, 70), mt_rand(0, 30),
    mt_rand(0, 70), mt_rand(0, 30));
$ImagickDraw->line(mt_rand(0, 70), mt_rand(0, 30),
    mt_rand(0, 70), mt_rand(0, 30));
/* Bild zeichnen, Format setzen und ausgeben*/
$Imagick->drawImage($ImagickDraw);
$Imagick->setImageFormat('jpeg');
session_start();
$_SESSION['captcha_phrase'] = $string;
echo $Imagick->getImageBlob();
$Imagick->destroy();
?>
```

HUBERT BENJAMIN RITZDORF

ist Student der Informatik an der Universität Bonn und arbeitet als Hilfskraft für die IT Research Division der NEC Europe Laboratories in Sankt Augustin.

Onlinequellen

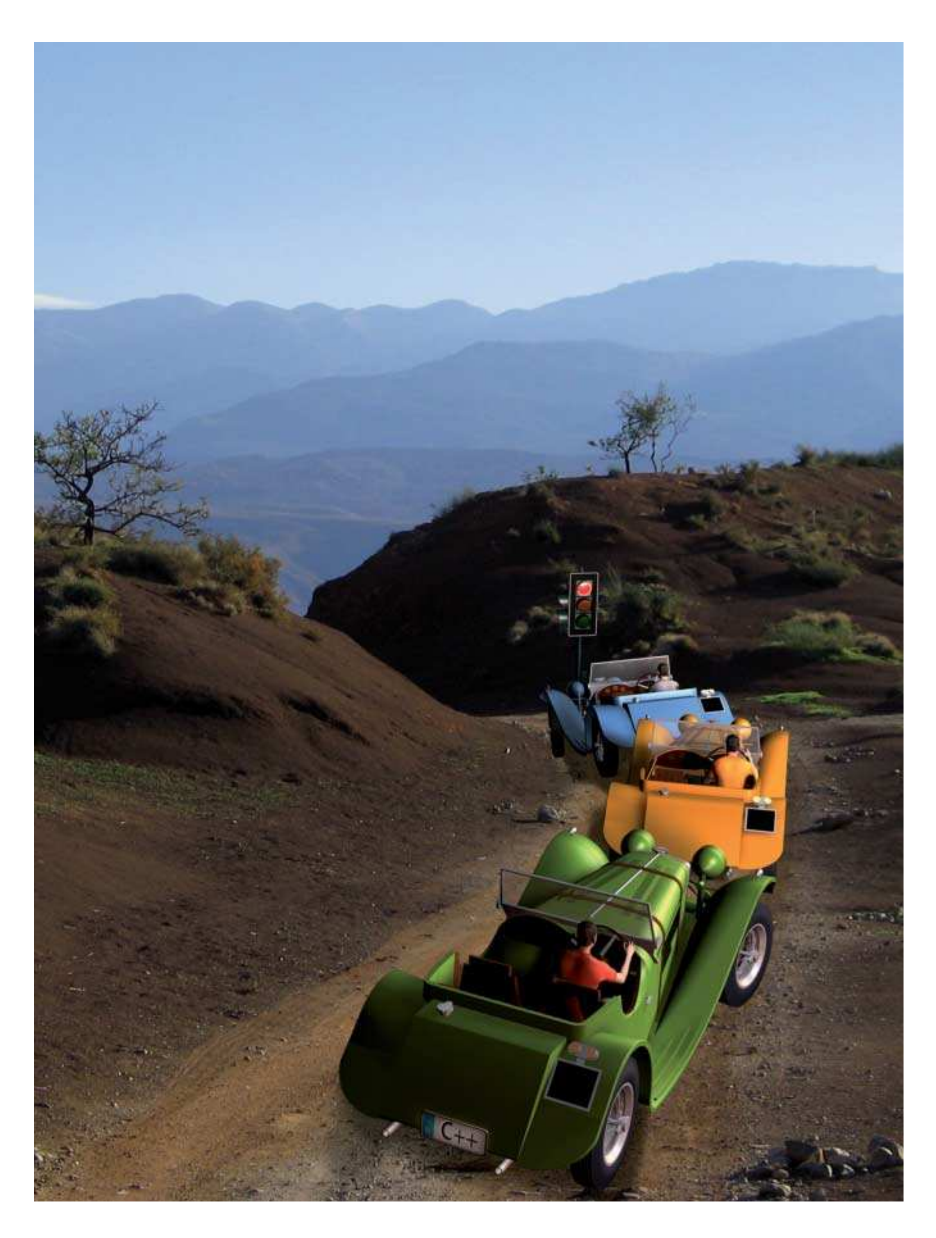
Imagick-Library	de.php.net/manual/en/ref.imagick.php
GD-Library	de.php.net/gd
Captcha-Breaking	www.puremango.co.uk/cm_breaking_captcha_115.php
Erkennen verzerrter Texte	research.microsoft.com/~kumarc/pubs/chellapilla_ceas05.pdf
Captcha Killer	www.captchakiller.com
AICaptcha	www.brains-n-brawn.com/default.aspx?vDir=aicaptcha
Breaking a Visual Captcha	www.cs.sfu.ca/~mori/research/gimpy/
PWNtcha	sam.zoy.org/pwntcha/
Alternativen zu Captchas	www.w3.org/2004/Talks/0319-csun-m3m
Captcha-Projekt	www.captcha.net/
PHP-Captcha-Objekt	pear.php.net/manual/en/package.text.text-captcha.php

Boost-Tutorial II: Thread- und Signal/Slot-Programmierung

Startsignale

Rüdiger Berlich

Die frei verfügbare Boost-Bibliothek lässt das Herz manches C++-Programmierers höherschlagen, da sie dringend benötigte und häufig vermisste Funktionen zur Verfügung stellt. Dieser zweite Tutorial-Teil spannt den Bogen von generalisierten Callback-Funktionen über Threads bis hin zur Signal/Slot-Programmierung und zum Umgang mit Zeit-Ausdrücken.



Thema des ersten Tutorial-Teils war die *Boost.Bind*-Bibliothek. Sie erlaubt die Zuordnung von Werten zu beliebigen Argumenten von Funktionen oder Funktionsobjekten sowie die automatische Erzeugung passender Funktionsobjekte. Die gemeinsame Verwendung von *Boost.Bind* mit der *Boost.Function*-Bibliothek (Header *<boost/function.hpp>*) bietet mehr als die Summe der Teile.

Boost.Function implementiert eine generalisierte Callback-Infrastruktur und ermöglicht es zudem, Funktionen und Funktionsobjekte wie Variablen zur späteren Verwendung zu speichern oder als Argument zu übergeben. Eine weitere Besonderheit besteht darin, dass der Benutzer nicht mehr zwischen Funktionsobjekten und -zeigern unterscheiden muss. Ein *boost::function*-Objekt mit einer bestimmten Signatur ist kompatibel mit allen funktionsartigen Entitäten mit derselben Signatur. Die Signatur ergibt sich aus dem Rückgabewert und den Argumenten.

Funktionen in Variablen speichern

Zur Erläuterung von *Boost.Function* soll die Klasse *functionTester* dienen (Listing 1). Alle Benutzerinteraktionen erfolgen über die beiden Methoden der Klasse – das Beispiel verzichtet daher der Einfachheit halber auf Konstruktor und Destruktor. Beide erzeugt der Compiler automatisch. Die Aufgabe der Klasse ist, mit einer gegebenen Methode einen String auszudrucken. Die Ausgabefunktion selber ist nicht Teil der Klasse, sondern wird mit der Methode *functionTester::setMessenger()* durch den Benutzer registriert. *setMessenger()* erhält ein *boost::function*-Objekt als Argument. Als Template-Argumente übergibt das Programm dem *function*-Objekt den gewünschten Rückgabewert (in diesem Fall einfach *void*) sowie ein oder mehrere Funktionsargumente (hier eine konstante Referenz auf einen *std::string*).

Bei *boost::function<void (const string&)> f* handelt es sich um ein herkömmliches Objekt, das man entsprechend behandeln kann. Das Beispielprogramm speichert es einfach in einer privaten Variable *f_* desselben Typs zur späteren Verwendung. Da das Programm keine Referenz übergeben hat, kopiert es das *boost::function*-Objekt. Der Umgang mit dieser Klasse ist also nicht viel komplizierter als etwa mit einem Integer-Wert.

Mitteilungen soll die *functionTester::printMessage()*-Funktion ausgeben. Sie erhält einen *std::string* als Argument und übergibt es an das gespeicherte *boost::function*-Objekt zur Auswertung. Natürlich muss hierbei eine gewisse Fehlerprüfung erfolgen – schließlich kann es sein, dass noch keine Funktion registriert wurde. Da *boost::function* implizit in einen Bool-Wert umgewandelt werden kann, ist dies einfach. Das Beispielprogramm ruft *f_(msg)* genau dann auf, wenn es in *f_* eine passende Funktion gespeichert hat. Ist dies nicht der Fall, gibt es eine Fehlermeldung aus.

Mit den passenden Funktionen und Funktionsobjekten lässt sich die Klasse *functionTester* ausprobieren. Eine einfache Funktion ohne Rückgabewert *void freeMessenger(const string& msg) {}* sowie eine Klasse *messenger* erfüllen diesen Zweck (Listing 2).

freeMessenger() sollte selbsterklärend sein. Die Klasse *messenger* implementiert mit *messenger::operator() (const string&)* ein Funktionsobjekt. Die Ausgabe des Strings erfolgt in der Memberfunktion *freeMessenger::messagePrinter()*.

Um zwischen dem Aufruf über den *operator()* und dem direkten Aufruf zu unterscheiden, erhält sie neben dem String ein zusätzliches Argument *int mode*, mit dem sie den jeweiligen Modus anzeigen kann.

Listing 3 zeigt, wie man die einzelnen Komponenten in *main()* zusammenbaut. Zunächst legt es einen *std::vector* von *shared_ptr*-Objekten an, die auf *functionTester*-Objekte zeigen. Diesen Container füllt es dann mit *functionTester*-Objekten. Die folgenden Ausführungen demonstrieren vier verschiedene Fälle.

Zur *shared_ptr*-Klasse hat bereits der erste Teil des Tutorials Erläuterungen gegeben. Es handelt sich hierbei um eine *Smartpointer*-Klasse aus Boost. Sie erspart es dem Entwickler, am Ende der Programmausführung mit *delete* die dynamisch allozierten Objekte zu löschen.

Anlässlich eines Leserkommentars (siehe Leserbriefe Seite 8) sei an dieser Stelle darauf hingewiesen, dass die durch den *shared_ptr* vorgenommene Allozierung von Objekten in der Form *shared_ptr<functionTester> p(new fun ctionTester);* anstelle von *functionTester *f = new functionTester();shared_ptr <functionTester> p(f);* erfolgen sollte (vgl. Listing 3 dieses Artikels; Listing 2 des ersten Tutorial-Teils handhabt dies anders). Der Grund dafür, die Semantik aus Listing 3 zu wählen, sind Ausnahmen, die ein Programm beispielsweise bei der Allozierung des Objekts werfen kann. Für weitere Regeln im Umgang mit *shared_ptr* sei auf den Abschnitt „Best Practices“ der Dokumentation verwiesen (siehe „Onlinequellen“ [b]).

functionTester::setMessenger() registriert einen Funktionszeiger auf die freie *freeMessenger*-Funktion. Es folgt ein Funktionsobjekt (*messenger()* erzeugt dabei ein temporäres Objekt der Klasse *messenger*). Im dritten Fall erzeugt *boost::bind* ein Funktionsobjekt, in dem ein direkter Aufruf von *functionTester::messagePrinter()* erfolgt. Damit die Memberfunktion dies weiß, muss der passende Wert *(2)* an das *mode*-Argument gebunden werden. Hier zeigt sich wieder, wie nützlich *boost::bind* ist. Dem vierten und letzten *functionTester*-Objekt wird bewusst keine Funktion zugewiesen.

Als letzte Aufgabe bleibt, die vier *printMessage()*-Funktionen der im Container gespeicherten *functionTester*-Objekte aufzurufen. Die Ausgabe des Programms sieht so aus:

```
Free messenger says: Hello World
Function object says: Hello World
Member function says: Hello World
No callback registered yet
```

Offensichtlich hat das Programm einen Funktionspointer und zwei unterschiedliche Funktionsobjekte erfolgreich als Argumente an *functionTester::setMessenger()* übergeben und in einer privaten Variablen der Klasse gespeichert. Dies ermöglicht Benutzern wie Bibliotheksdesignern eine bislang nicht dagewesene Freiheit.

Im vierten *functionTester*-Objekt ist keine Funktion gespeichert. *printMessage()* fängt diesen Fehler ab, indem es die Fähigkeit des *boost::function*-Objekts nutzt, sich in einen Bool-Wert wandeln zu lassen. Hier zeigt

Listing 1: Die Klasse *functionTester*

```
class functionTester
{
public:
    void printMessage(const string& msg){
     if(f_)
         f_(msg);
     else
         cout << "No callback registered yet" << endl;
    }

    void setMessenger(boost::function<void (const string&)> f){
     f_ = f;
    }
private:
    boost::function<void (const string&)> f_;
};
```

Listing 2: *freeMessenger*-Funktion und *messenger*-Klasse

```
void freeMessenger(const string& msg){
    cout << "Free messenger says: " << msg << endl;
}
class messenger
{
public:
    void operator()(const string& msg){
     messagePrinter(msg, 1);
    }
    void messagePrinter(const string& msg, int mode){
     if(mode==1)
         cout << "Function object says: " << msg << endl;
     else
         cout << "Member function says: " << msg << endl;
    }
};
```

Listing 3: *main()*

```
int main(int argc, char **argv){
    typedef vector<shared_ptr<functionTester> > ftType;
    ftType ftContainer;
    for(unsigned int i=0; i<4; ++i){
     shared_ptr<functionTester> p(new functionTester);
     ftContainer.push_back(p);
    }
    // Registering a function pointer
    ftContainer[0]->setMessenger(&freeMessenger);
    // Registering a function object
    ftContainer[1]->setMessenger(messenger());
    // Registering a member function with partial argument binding
    messenger Messenger;
    ftContainer[2]->setMessenger(
     boost::bind(&messenger::messagePrinter,
             &Messenger,_1, 2)
     );
    // Position 3 intentionally left empty
    ftType::iterator it;
    for(it=ftContainer.begin(); it!=ftContainer.end(); ++it)
     (*it)->printMessage("Hello World");
    return 0;
}
```

iX-TRACT

- Die *Boost.Function*-Bibliothek erleichtert C++-Programmierern den Umgang mit Callbacks, unter anderem dadurch, dass sie nicht mehr zwischen Funktionsobjekten und Zeigern unterscheiden müssen.
- *Boost.Thread* profitiert von *Boost.Function* und trägt dafür Sorge, dass mehrere Threads in einem Programm problemlos parallel laufen können.
- Auch die *Boost.Signals*-Bibliothek setzt auf die Konzepte von *Boost.Function*. Der Signal/Slot-Mechanismus erlaubt es, Funktionen einer Klasse aus einer anderen heraus aufzurufen, ohne dass der Aufrufer wissen muss, welche Funktion sein Signal aktiviert.

sich wieder das durchdachte Design der Bibliothek.

Boost.Bind und *Boost.Function* besitzen noch eine Schwesterbibliothek: *Boost.Lambda*. Sie soll hier nur gestreift werden. Die aus Lisp und Python bekannte Technik der anonymen Funktionen (auch Lambda-Funktionen genannt) erlaubt es, am Ort eines Aufrufs Funktionen ohne Namen zu erstellen. Wenn Funktionen nur an einer Stelle im Programmtext vorkommen, kann man so die Komplexität einer Anwendung reduzieren. Boost implementiert diese Möglichkeit nun auch für C++. Bemerkenswert ist, dass *Boost.Lambda* ohne Änderungen des C++-Standards auskommt. Der Artikel wird im Folgenden noch mehrfach auf *Boost.Function* und *Boost.Bind* eingehen, *Boost.Lambda* kommt jedoch nicht weiter zur Sprache.

Eine Bibliothek, die von *Boost.Function* stark profitiert, ist *Boost.Thread*. Threads sind parallele Ausführungseinheiten in einem Programm. Durch die parallele Nutzung etwa verschiedener Prozessoren steigt (im Idealfall) die Ausführungsgeschwindigkeit.

Listing 4: Programm mit einem Thread

```
unsigned int global_store = 0;
const unsigned int MAXCOUNTER = 500;
void counter(unsigned int n = 1){
  while(true){
    if(global_store > MAXCOUNTER) break;
    if(++global_store%100 == 0)
      cout << "In counter " << n << ": " << global_store << endl;
  }
}
int main(int argc, char **argv){
  boost::thread cntThrd1(&counter);
  cntThrd1.join();
  return 0;
}
```

Listing 5:
Erweiterte *counter()*-Funktion

```
boost::mutex cnt_mutex;
unsigned int global_store = 0;
const unsigned int MAXCOUNTER = 1000;
void counter(unsigned int n = 1){
  unsigned int tmp_global_store;
  while(true){
    { // additional scope used for scoped_lock
      boost::mutex::scoped_lock lock(cnt_mutex);
      if(global_store >= MAXCOUNTER) break;
      if(++global_store % 100 == 0)
     cout << "In counter function " << n << ": " \
                                         << global_store << endl;

      tmp_global_store = global_store;
      // save variable for later usage
    }

    // Put the thread to sleep for random amount of time
    MTRand rnd(tmp_global_store);
    // initialise with current global store
    xtime xt;
    boost::xtime_get(&xt, boost::TIME_UTC);
    // the current time
    xt.sec  += 0; // seconds
    xt.nsec += rnd.randInt(1000000);
    // random amount of nanoseconds in range [0,1000000]
    boost::thread::sleep(xt);
  }
}
```

Bekannte, portable Thread-Implementierungen sind beispielsweise die Posix Threads (mit einer C-API) oder die Thread-Bibliothek von Trolltechs Qt. Erstere ist unter C++ etwas beschwerlich zu nutzen. Callbacks werden über Funktionspointer realisiert, und das Design kann (aufgrund der Einschränkungen der Programmiersprache C) nicht der unter C++ eigentlich obligatorischen Typsicherheit folgen. Qts Thread-Bibliothek ist recht komfortabel und auch für freie Programme einsetzbar (sofern diese unter der GPL stehen). Nicht jeder mag allerdings seine eigenen Programme ebenfalls unter die GPL stellen oder alternativ eine Qt-Lizenz erwerben. Hier kommt das ebenfalls portable Boost ins Spiel.

Ein Thread muss die Funktion kennen, die er nach seinem Start ausführen soll. In Qt leitet man hierzu eine Klasse von *QThread* ab und implementiert eine *run()*-Methode. Obwohl einsichtig, schränkt dieses Design etwas ein. Boost verlässt sich hier voll auf die *Boost.Function*-Bibliothek und erscheint somit flexibler. Die auszuführenden Methoden werden nach Wahl als Funktionsobjekt oder als -pointer übergeben. *Boost.Function* kommt mit beiden zurecht. Einen einfachen Fall zeigt Listing 4.

Eine Funktion *counter()* inkrementiert eine globale Variable *global_store*, bis diese einen Maximalwert erreicht hat. Alle 100 Inkrementierungen gibt sie den aktuellen Status aus. Das Beispiel startet nur einen Thread in *main()* (genau genommen zwei, da *main()* weiter aktiv bleibt). Um später verschiedene Threads unterscheiden zu können, erhält *counter()* ein Argument *n* mit dem Standardwert *1*.

Der Start des Threads erfolgt durch die Instanziierung eines *boost::thread*-Objekts, dem ein Funktionszeiger auf *counter()* übergeben wird. Um das Funktionsargument muss sich der Programmierer wegen des Standardwerts von *1* an dieser Stelle nicht kümmern.

main() wartet mit *boost::thread::join()* auf das Ende der Ausführung von *counter()*. Ist dies erreicht, beendet sich das Programm. Das Beispiel ist zugegebenermaßen noch nicht sehr aufregend. Es soll jetzt auf mehrere Threads erweitert werden. Dies betrifft zunächst die *counter()*-Funktion (Listing 5).

Hier wird es schon interessanter. Das Programm soll mehr als eine *counter()*-Funktion gleichzeitig betreiben. Es gibt jedoch eine Reihe von Ressourcen, die von mehreren Funktionen gemeinsam genutzt werden. Offensichtlich ist dies bei der *global_store*-Variable, die alle Funktionen schreiben sollen. Diesen Zugriff gilt es zu synchronisieren, sonst herrscht Chaos.

Den Zugriff reguliert man deshalb oft mit einem Mutex. Er muss vor jedem Zugriff auf eine gemeinsame Ressource blockiert werden (*Boost.Thread* sorgt dafür, dass dies sicher vor verschiedenen Ausführungssträngen erfolgen kann). Bei blockiertem Mutex hält die Ausführung eines Threads an. Erst wenn der Mutex wieder freigegeben wurde und der Thread ihn für sich beanspruchen konnte (andere Threads können um diesen Mutex ebenfalls konkurrieren), kann seine Ausführung weitergehen. Halten sich alle Threads an die Konvention, vor dem Zugriff auf eine gemeinsame Ressource einen Lock auf einen Mutex zu setzen, ist die Gefahr gebannt. Dabei sollte allerdings klar sein, dass das Anhalten eines Threads nicht gerade zur Beschleunigung eines Programms beiträgt.

Das Beste aus parallelen Welten

Nicht nur gemeinsame Variablen gilt es zu schützen. Auch Funktionsaufrufe können kritisch sein, wenn diese nicht auf die Verwendung mit Threads zugeschnitten sind. Die geänderte *counter()*-Funktion macht dies offensichtlich.

Innerhalb der *while()*-Schleife versucht das Programm zunächst, den Zugriff auf den Mutex *cnt_mutex* zu erhalten. Wie in Boost üblich, bedient sich das Beispiel dabei der Technik, automatische Variablen die Freigabe einer Ressource übernehmen zu lassen. In diesem Fall gibt das *scoped_lock*-Objekt automatisch den Mutex frei, wenn es das Ende seiner Lebensdauer erreicht. Um diesen Prozess zu steuern, rahmen geschweifte Klammern den entsprechenden Bereich ein. So entsteht ein eigener Geltungsbereich.

Innerhalb dieses Bereichs prüft das Programm – nachdem es den Zugriff auf den Mutex erlangt hat – zunächst, ob die Zählvariable ihren Maximalwert überschritten hat. Hier zeigt sich übrigens, warum in diesem Fall eine Endlosschleife sinnvoll ist. Eine Abfrage *while(global_store < MAXCOUNTER)* wäre zwar möglich (das Lesen von *global_store* außerhalb des Mutex-Schutzes ist unkritisch, da es sich um eine atomare Operation handelt). Jedoch wird der Mutex erst danach in Anspruch genom-

men. Es gibt also keine Sicherheit, dass beim Zugriff auf *global_store* nicht ein anderer Thread dessen Wert bereits wieder geändert hat. Alle Schreib- und Lesezugriffe auf die gemeinsam genutzten Ressourcen sollten daher im Schutz des Mutex erfolgen. Aus demselben Grund speichert *counter()* den aktuellen Wert von *global_store* in einer lokalen Variablen zur späteren Verwendung.

Der gleichzeitige Zugriff auf *std::cout* aus mehreren Threads gleichzeitig kann wegen des Caching übrigens Probleme verursachen. Auch hier ist der Schutz eines gesperrten Mutex hilfreich.

Mit dieser Infrastruktur lässt sich *global_store* nun schon aus mehreren Threads gleichzeitig hochzählen. Allerdings ginge dies bei einem Maximalwert von 1000 zu schnell – wahrscheinlich kämen nicht alle Threads zum Zuge.

Daher wird festgelegt, dass der Thread nach der Freigabe des Mutex eine Weile inaktiv ist. Dies würde man normalerweise mit der *usleep()*-Funktion realisieren. Im Regelfall ist sie aber nicht für die Benutzung in Threads geeignet. Deshalb kommt hier die *Boost.Thread*-eigene *sleep*-Funktion zum Einsatz. Die Berechnung einer Zufallszahl soll verhindern, dass die Threads immer gleich lang schlafen. Hier zeigt sich wieder, dass man bei der Thread-Programmierung meist nicht einfach eine der Standardfunktionen der C-Bibliothek verwenden kann, da diese beispielsweise gelegentlich einen internen Status speichern. Die gemeinsame Verwendung aus mehreren Threads kann dann zu Fehlern führen. Der Einfachheit halber verwendet das Beispielprogramm eine frei verfügbare Implementierung von Zufallszahlen: die Klasse *MTRand* aus *MersenneTwister.h* von Richard J. Wagner [c].

Da ein lokales Objekt erzeugt wird, das keine globalen Variablen verwendet, ist dies unkritisch. Als Eingabe für den Zufallszahlengenerator dient der zwischengespeicherte Wert von *global_store*. *MTRand::randInt(1000000)* liefert einen Integer-Wert im Bereich [0,1000000], der für die zufällige Wartezeit verwendet werden kann. Leider ist der Umgang mit *boost::thread::sleep()* etwas umständlich.

Der Start der Threads erfolgt erneut in *main()* (Listing 6). Im Unterschied zum ersten Beispiel verwendet dieses eine *thread_group* und fügt diesem Container für Thread-Objekte auf drei verschiedene Weisen Threads hinzu. Auch hier zeigen sich wieder die Vorteile von *Boost.Function*.

Der erste Thread verwendet einen Funktionspointer auf *counter()*. Im zweiten Fall kommt ein Funktionsobjekt der Klasse *counter_class* zum Einsatz. Und schließlich wird mit *boost::bind* ein *boost::function*-Objekt initialisiert, das dem dritten Thread als Argument übergeben wird. Die Fälle lassen sich durch Indizes unterscheiden, die Argumente der *counter()*-Funktion sind. Ein Ausschnitt aus der Ausgabe des Programms sieht aus wie folgt:

```
In counter function 2: 600
In counter function 3: 700
In counter function 2: 800
In counter function 1: 900
```

Es ist übrigens sinnvoll zu überprüfen, ob überhaupt eine Thread-Implementierung in Boost zur Verfügung steht. Dies kann im Kopfteil des Programmtexts geschehen. Wann immer Threads in Boost verfügbar sind, ist *BOOST_HAS_THREADS* definiert. Man kann also einfach schreiben:

```
#ifndef BOOST_HAS_THREADS
#error "Error: No thread support
#enÏdif
```

Den Rest erledigt in diesem Fall der Präprozessor.

Selbst dieses einfache Beispiel sollte deutlich gemacht haben, dass Thread-Programmierung alles andere als eine

Bewegliche Ziele

Tutorials beschreiben oft bewegliche Ziele. Boost (und die für dieses Tutorial verwendete Plattform) fügt sich in diese Regel ein. Seit der Veröffentlichung des ersten Teils dieser dreiteiligen Serie wurde nicht nur eine neue Version von Opensuse freigegeben. Mit Boost 1.34.1 steht mittlerweile eine Plattform zur Verfügung, die in puncto Funktionsvielfalt und Fehlerfreiheit gegenüber der Version 1.33.1 deutliche Vorteile bietet. Da Opensuse 10.3 weiterhin mit dem „alten“ Boost 1.33.1 ausgeliefert wird, soll hier kurz auf die Übersetzung der Bibliothekssammlung in diesem Umfeld eingegangen werden.

Unter Linux ist der eigentliche Übersetzungsvorgang fast trivial – wie bei den meisten Open-Source-Paketen beschränkt er sich auf den Aufruf von *configure*, gefolgt von *make*. Die jeweils neuesten Quellen sind auf der Boost-Website [a] verfügbar. Etwas komplizierter gestaltet es sich, wenn man zusätzliche Funktionen wünscht, die im Standardpaket nicht enthalten sind. Neben den Basiskomponenten gibt es eine Vielzahl an Bibliotheken, die dem Boost-Entwicklungsmodell folgen und, was wichtiger ist, sich in dessen Infrastruktur einfügen. Ein Beispiel ist die in diesem Artikel erwähnte *threadpool*-Bibliothek von Philipp Henkel. Daneben gibt es Libraries, die zwar offiziell Eingang in Boost gefunden haben, aber noch nicht in der aktuellen Release enthalten sind. Besonders wichtig ist hier die *asio*-Bibliothek, die Boost um Funktionen für asynchrones IO, insbesondere auf Netzwerkebene, ergänzt. Auf *asio* geht der dritte Teil des Tutorials näher ein. Für die Verwendung von *asio* und *threadpool* ist unter Boost 1.34.1 etwas Handarbeit nötig.

Der erste Schritt ist das Herunterladen und Entpacken der Boost-Quellen – in diesem Fall der Version 1.34.1. Die *threadpool*-Bibliothek ist in der Version 0.2.4 auf der *threadpool*-Website [f] zu finden. *asio* ist in der Version 0.3.8 ebenfalls über Sourceforge erhältlich [h]. Beim Download sollte das mit „boost“ beginnende Paket gewählt werden. *asio* hat seinen Ursprung außerhalb von Boost, und so gibt es eine weitere Version der Bibliothek, die die Boost-Konventionen nicht beachtet. Beide Pakete entpackt man in einem beliebigen Verzeichnis.

Den Inhalt des *asio*-Verzeichnisses kopiert der Befehl

```
cp -ir * ../boost_1_34_1/
```

in das Boost-Verzeichnis (dabei wird offensichtlich angenommen, dass sich dieses im Elternverzeichnis befindet). Bei der *threadpool*-Bibliothek wechselt man in das Unterverzeichnis *threadpool/include* des Archivs und kopiert den Inhalt mit

```
cp -ir * ../../../boost_1_34_1/boost/
```

in das Unterverzeichnis *boost* des Boost-Archivs. Auch hier gelten dieselben Annahmen über den Ort der Boost-Bibliothek.

Unter Opensuse 10.3 gibt es eine weitere Besonderheit. Durch ein Problem im Zusammenspiel mit Opensuses Standardcompiler G++ 4.2.1 und Boosts Build-System *bjam* ist es nötig, in einem Skript von Hand eine Compiler-Einstellung vorzunehmen. Hierzu ändert man Zeile 155 in der Datei *tools/jam/src/build.jam* (relativ zur Archivwurzel), sodass dort der Text

```
toolset gcc gcc : "-o " : -D
    : -pedantic -fno-strict-aliasing
```

steht. Nun reicht das beschriebene *configure; make; make install* für die Übersetzung der Bibliothek aus.

Während *asio* in der kommenden Boost-Version 1.35 enthalten sein soll, dürfte es bei *threadpool* noch etwas länger dauern, bis sie Eingang in die Standarddistribution erhält.

Man sollte sich übrigens nicht von der niedrigen Versionsnummer schrecken lassen – beide Bibliotheken sind stabil.

triviale Angelegenheit ist. Wer sich dennoch bislang nicht hat abschrecken lassen, kann über die Boost-Webseite [a] und ein Online-Tutorial [d] weitere Anregungen zur Thread-Programmierung mit Boost sammeln. Die Webseiten bieten darüber hinaus eine Reihe von Hilfestellungen zu diesem Thema.

Besonders erwähnenswert erscheint die *threadpool*-Bibliothek von Philipp Henkel [f]. Sie ist auf die Verwendung mit Boost abgestimmt, hat aber noch keinen Eingang in die Kernsammlung gefunden. Dem Namen entsprechend implementiert *threadpool* das Thread Pool Pattern. Die Aufgabe von *threadpool* kommt damit der eines Vermittlers (Broker, Dispatcher) gleich: Hat man eine Reihe M an Aufgaben, die durch eine Anzahl N von Threads abgearbeitet werden sollen, so weist *threadpool* den Threads ihre jeweiligen Aufgaben zu. Hat ein Thread seine Aufgabe beendet, erhält er eine neue aus der Liste noch unbearbeiteter Tasks. Die Anzahl der Aufgaben ist typischerweise deutlich größer als die der Threads.

Je nach Implementierung bedeutet das Erzeugen und Zerstören von Threads einen signifikanten Aufwand. Mit *threadpool* kann man das vermeiden. Für die Bibliothek bedeutet das zwar einigen Verwaltungsaufwand, allerdings sieht der Benutzer nichts von der Komplexität dieses Vorgehens. Er weist *threadpool* lediglich Aufgaben zu und überlässt der Bibliothek die Verwaltungsarbeit. Der Kasten „Bewegliche Ziele“ enthält einige weitere Details zu *threadpool*.

Das Beispiel aus Listing 5 und 6 hätte man mit einem Thread-sicheren Signal/Slot-Mechanismus möglicherweise einfacher implementieren können. Ein solcher steht unter Boost leider (noch) nicht zur Verfügung (auch wenn die *Vault*-Bibliotheken bereits einen Prototypen enthalten). Ist man jedoch bereit, auf Threads zu verzichten, bietet die *Boost.Signals*-Bibliothek von Douglas Gregor eine ausgereifte Implementierung des Signal/Slot-Konzepts an. Wiederum kommen dabei die Konzepte von *Boost.Function* ins Spiel.

Signale setzen in gutem Glauben

Bei dem Signal/Slot-Mechanismus handelt es sich um eine weitere Möglichkeit, Funktionen einer Klasse aus einer anderen heraus aufzurufen (beschrieben als „Senden eines Signals“). Der Aufruf kann auch die Übergabe von Parametern und die Rückgabe von Werten beinhalten.

Das ist an sich nicht revolutionär. Bedeutung erlangt es dadurch, dass der Aufrufer nicht darüber Bescheid wissen muss, welche Funktion sein Signal aktiviert. Das bestimmt der Programmierer außerhalb der aufrufenden und ausführenden Klasse, indem er Signale mit Slots verbindet. Dabei ist es erlaubt, einem Signal mehrere Slots zuzuweisen, sodass ein Signal mehr als eine Zielfunktion aktiviert.

Bekannt geworden ist dieser Mechanismus vermutlich durch Trolltechs Implementierung in der Qt-Bibliothek, die unter anderem der KDE-Oberfläche von Linux zugrunde liegt. Trolltech bemüht hierzu Makros und den Meta Object Compiler *moc* zusammen mit eigenen Schlüsselwörtern. Dies kommt fast einer Erweiterung des C++-Sprachstandards gleich. Boost kommt ohne solche Umwege aus. Umgekehrt ist aber zu erwähnen, dass Qt in der aktuellen Version durchaus Signal/Slot-Verbindungen zwischen Threads zulässt und (nicht nur) damit wohl flexibler ist.

Ziel dieses Abschnitts ist es, eine Art Time Server zu erstellen. Über dasselbe Signal soll er – je nach Slot – die aktuelle Zeit oder die seit dem 1.1.2001 vergangene ausgeben. Um dies zu bewältigen, zunächst ein Blick auf eine weitere Boost-Komponente – Jeff Garlands *date_time*-Bibliothek.

Ein einfaches Beispiel demonstriert das Vorgehen (Listing 7, Header: *<boost/date_time.hpp>*). Zunächst fragt das Programm in Abständen von 0,1 und 0,2 Sekunden nach der aktuellen Zeit, mit einer Auflösung im Mikrosekundenbereich. Die Werte speichert es in *ptime*-Objekte (Posix Time). Aus ihnen kann man intuitiv die vergangene Zeit berechnen, etwa durch den Aufruf *time_duration tx = (t2-t1);*.

time_duration-Objekte lassen sich auch miteinander vergleichen. *tx* und *ty* sind beides *time_duration*-Objekte. Der Vergleich *tx >= ty* erlaubt eine Aussage darüber, ob *tx* größer (oder eventuell gleich) *ty* ist. Die Ausgabe kann in verschiedener Granularität erfolgen, etwa in Stunden, Minuten oder Sekunden. Umgekehrt lässt sich ein *time_duration*-Objekt zum Beispiel mit Stunden, Minuten oder Sekunden initialisieren.

Die Ausgabe von Listing 7 lautet:

```
Test verschiedener Zeitabstaende
tx=00:00:00.101221, ty=00:00:00.201106
ty ist groesser
Stunden seit dem 1.1.2001: 60074
Zeitintervall aus Stunden, Minuten, Sekunden:
h:m:s = 17:22:3
Gesamt-Sekunden = 62523
Initialisierung mit String "00:00:00.000"
total_seconds = 0
```

Die Verwendung von *date_time* ist so intuitiv und die Dokumentation auf der

Listing 6: Start der Threads

```
class counter_class{
public:
  void operator()(){ counter(2); }
};

int main(int argc, char **argv){
  boost::thread_group thrd_grp;
  counter_class cnt_class;
  boost::function<void (void)> bf_counter =
    boost::bind(counter, 3);

  thrd_grp.create_thread(&counter);
  thrd_grp.create_thread(cnt_class);
  thrd_grp.create_thread(bf_counter);

  thrd_grp.join_all();
  return 0;
}
```

Listing 7: *date_time*

```
int main(int argc, char **argv){
  cout << "Test verschiedener Zeitabstaende" << endl;
  ptime t1 = microsec_clock::local_time();
  usleep(100000);
  ptime t2 = microsec_clock::local_time();

  ptime t3 = microsec_clock::local_time();
  usleep(200000);
  ptime t4 = microsec_clock::local_time();

  time_duration tx = (t2-t1);
  time_duration ty = (t4-t3);
  cout << "tx=" << tx << ", ty=" << ty << endl;

  if(tx >= ty) cout << "tx ist groesser" << endl;
  else cout << "ty ist groesser" << endl;

  ptime t5(date(2001, Jan, 1));
  ptime t6(second_clock::local_time());
  cout << "Stunden seit dem 1.1.2001: " << (t6-t5).hours() << endl;

  cout << "Zeitintervall aus Stunden, Minuten, Sekunden:" << endl;
  time_duration td = hours(17) + minutes(22) + seconds(3);
  cout << "h:m:s = " << td.hours() << ":" << td.minutes() << ":" << td.seconds() << endl;
  cout << "Gesamt-Sekunden = " << td.total_seconds() << endl;

  cout << "Initialisierung mit String \"00:00:00.000\"" << endl;
  time_duration empty(duration_from_string("00:00:00.000"));
  cout << "total_seconds = " << empty.total_seconds() << endl;
}
```

Tutorialinhalt

Listing 8: Zwei Time Server

```
class total_time_server{
public:
  string operator()(void){
    ptime t1(second_clock::local_time());
    return to_simple_string(t1);
  }
};
class time_difference_server{
public:
  string operator()(string startDate){
    ptime t1(from_simple_string(startDate));
    ptime t2(second_clock::local_time());
    time_duration td = t2-t1;
    return to_simple_string(td);
  }
};
class time_retriever
{
public:
  boost::signal<string (void)> getTime;
  boost::signal<string (void),
    aggregate_values\<vector<string> > > getAllTimes;
};
```

Listing 9: *main()* des Signal-Beispiels

```
int main(int argc, char **argv){
  total_time_server tts;
  time_difference_server tds;
  time_retriever tr;
  boost::signals::connection c=tr.getTime.connect(tts);
  cout << tr.getTime() << endl;
  c.disconnect();
  c=tr.getTime.connect(boost::bind<string>(tds,"2001-01-01"));
  cout << tr.getTime() << endl;
  c.disconnect();
  if(!c.connected()) cout << "Connection terminated!" << endl;
  boost::signals::connection c1=tr.getAllTimes.connect(1,tts);
  boost::signals::connection c2=
    tr.getAllTimes.connect(0,boost::bind<string>(tds,"2001-01-01"));
  vector<string> results = tr.getAllTimes();
  copy(results.begin(), results.end(),
       ostream_iterator<string>(cout, " "));
  cout << endl;
  return 0;
}
```

Boost-Homepage so gut, dass keine weiteren Erläuterungen zu dieser Bibliothek notwendig sein dürften. Daher zurück zum Beispielprogramm.

Zunächst führt es drei kleine Klassen ein: *total_time_server*, *time_difference_server* und *time_retriever* (Listing 8). Die ersten beiden Klassen implementieren Funktionsobjekte, die bei Aufruf einen *std:string* mit einem Zeitausdruck zurückgeben. Im ersten Fall ist dies die aktuelle Zeit, im zweiten die seit einem als Argument übergebenen Startzeitpunkt vergangene Zeit. Die Funktionsweise beider Klassen sollte nach der Einführung der *date_time*-Bibliothek verständlich sein.

time_retriever enthält zwei Signale. Das erste – *getTime* – ist zur Speicherung eines einzelnen Slots (also der Zielfunktion) gedacht. Die Syntax sollte nach der Einführung in *Boost.Function* zu Beginn dieses Artikels klar sein.

Das zweite Signal ist für die gleichzeitige Speicherung verschiedener Slots gedacht. Beim Aufruf dieses Signals stellt sich die Frage, wie man die Rückgabewerte der Slots verwenden soll. Ohne weitere Festlegung ist das der Rückgabewert des zuletzt aufgerufenen Slots. Wer will, kann über einen weiteren Template-Parameter im Signal *getAllTimes* selbst festlegen, wie die Rückgabewerte zu verwenden sind. Im vorliegenden Fall sollen alle Rückgabewerte (es handelt sich ja um Strings) in einem *vector<string>* gespeichert werden.

Die Implementierung von *aggregate_values<>* ist der Dokumentation zu *Boost.Signals* entlehnt und soll hier nicht weiter besprochen werden. Interessenten finden weitere Informationen auf der Boost-Website [g]. Hier geht es jetzt mit der *main()*-Funktion des Signal-Beispiels weiter (Listing 9).

Zunächst erfolgt die Instanziierung der Objekte der Signal- und Slot-Klassen. Der Aufruf *connect()* des Signals verbindet Signal und Slot. Im ersten Fall ist dies einfach, da *total_time_server ::operator()* keine Argumente erwartet. Das Signal *getTime* des *time_retriever*-Objekts kann man nun wie eine normale Memberfunktion aufrufen.

Der Aufruf von *disconnect()* trennt die Verbindung zwischen Signal und Slot wieder. Stattdessen wird der *time_difference_server* mit dem *getTime*-Signal verbunden. Da der *time_difference _server::operator()* ein Argument erwartet (den Offset für die Berechnung der Zeitspanne), muss erneut *boost::bind* aushelfen. Das Vorgehen sollte klar sein: *boost::bind* erzeugt selbst wieder ein Funktionsobjekt, das *time_difference _server::operator()* mit dem passenden Argument aufruft. Was nicht passt, wird passend gemacht.

Nach dem erneuten Trennen der Verbindung überprüft der *connection::connected()*-Aufruf, ob noch eine Verbindung existiert. Nach demselben Muster werden beide Slots gleichzeitig mit dem *time_retriever::getAllTimes*-Signal verbunden. Wie gehabt ist die Aktivierung des Signals identisch mit einem normalen Funktionsaufruf. Anders als zu Beginn ist das Ergebnis aber die Ausführung beider Funktionsobjekte. Dank *aggregate_values* ist das Resultat beider Aufrufe in einem *vector<string>* gespeichert. Dessen Inhalt wird zur Kontrolle ausgeben. Die Ausgabe von Listing 9 sieht etwa aus wie folgt:

```
2007-Nov-09 02:35:50
60074:35:50
Connection terminated!
60074:35:50 2007-Nov-09 02:35:50
```

Hier ist zunächst der separate Aufruf beider Slots ersichtlich, gefolgt von der Trennung der Verbindung. Es folgt der (fast) gleichzeitige Aufruf beider Slots über *getAllTimes()*. Auch hier lässt sich *Boost.Signals* intutitiv verwenden.

Ausblick

Neben all dem Lob der Bibliothek soll ein Problem nicht verschwiegen werden. Viele Klassen und Funktionen sind auf Templates aufgebaut, die tief ineinander verschachtelt sind. Selbst kleine Fehler führen so zu Meldungen, für deren Interpretation Erfahrung hilfreich ist. Oft erstreckt sich eine Meldung über mehrere Zeilen, und von diesen Fehlermeldungen kann es Hunderte auf einmal geben. Ähnliches gilt für die Fehlersuche – Template-Klassen geben sich alles andere als zugänglich in einem Debugger.

Allerdings ist dies kein spezifisches Boost-Problem, sondern dasselbe gilt für die STL. Manche Boost-Bibliotheken versuchen jedoch, durch kleine „Fallen“ Fehler bereits zur Übersetzungszeit zu identifizieren. Und das führt zu mehr Meldungen der genannten Art.

Der dritte Teil des Boost-Tutorials wird einen Schwerpunkt auf die Serialisierung von Objekten und auf die Netzwerkkommunikation legen. Boost stellt hierfür Funktionen zur Verfügung, die man in der C-Famile sonst nur von Sprachen wie Java und C# kennt. (ka)

DR. RÜDIGER BERLICH

führt am Institut für Wissenschaftliches Rechnen des Karlsruhe Institute of Technology (KIT) eine Ausgründung aus dem Bereich der Parameteroptimierung in verteilten Umgebungen durch.

Onlinequellen

[a] Boost-Homepage	www.boost.org
[b] Dokumentation zu *shared_ptr*	www.boost.org/libs/smart_ptr/shared_ptr.htm
[c] *MersenneTwister.h* von Richard J. Wagner	www-personal.umich.edu/~wagnerr/MersenneTwister.h
[d] Bill Kempf; The Boost.Threads Library	www.ddj.com/cpp/184401518
[e] C++ Multithreading Tutorial	paulbridger.net/multithreading_tutorial
[f] *threadpool*-Bibliothek	threadpool.sf.net
[g] Dokumentation zu *Boost.Signals*	www.boost.org/doc/html/signals.html
[h] *asio*-Bibliothek	asio.sourceforge.net

Komprimieren und Tunnel bauen mit *ssh*

Vielseitig

Susanne Nolte

Verschlüsseln, authentifizieren, komprimieren – mit solchen Fähigkeiten ausgestattet ist die SSH mehr als nur eine sichere Remote-Shell.

Die Secure Shell oder *ssh* gehört schon seit Längerem zum Standardwerkzeug für das Arbeiten auf entfernten Rechnern. Doch ihre Fähigkeiten gehen wesentlich weiter: Zum einen beschränken sich ihre Tunnelfähigkeiten nicht auf X11-Sessions, zum anderen kann sie den Netzverkehr auf das Wesentliche reduzieren.

Denn mit der Option *–C* auf die Reise geschickt, komprimiert die SSH alle Daten mit dem *gzip*-Algorithmus. Allerdings sollte man hier auf die Ressourcen achten: Während die SSH-Kompression den Verkehr auf langsamen Verbindungen massiv beschleunigen kann – vor allem, wenn sie Grafikdaten überträgt –, kann das in schnellen Netzen das Gegenteil bewirken, wenn nämlich die CPUs als Bottleneck sich erweisen.

Ebenfalls lohnen kann sich das Komprimieren beim Kopieren ganzer Verzeichnisbäume – einfach per Pipe durch die SSH geschickt etwa mit *tar cf – <localdir> | ssh –C <remotehost> „cd <remotedir> && tar xpf –“*.

Einige Dämonen wie *fetchmail* oder *exim* besitzen SSH-Plug-ins, mit denen sich die Verbindung verschlüsseln lässt. *rsync* arbeitet inzwischen standardmäßig mit einem SSH-Tunnel. Versionen, die das nicht tun, sollten sich mit der Option *–e ssh* oder *––rsh=ssh* zum Pipen durch einen SSH-Tunnel bewegen lassen.

Eine Alternative zur Pipe bietet das Local Port Forwarding. Dafür etabliert man einen Tunnel zwischen einem freien lokalen Port und dem entfernten Zielport nach dem Muster *ssh –L <localport>:<targethost>:<targetport> <user>@<authhost>*. Um etwa eine Web-Session, die nicht über HTTPS angeboten wird, zu verschlüsseln, genügt der Befehl *ssh –L 6080:<webserver>:80 <user>@<sshserver>* in einem Terminal, danach lädt man mit dem Browser *http://localhost:6080*. Das Gleiche gilt für ungesicherte IMAP-, POP3-, SMTP- oder VNC-Sessions: Nach dem Etablieren des Tunnels ist in der Client-Software lediglich *localhost* samt gewähltem Port als Server einzutragen. Dabei müssen Anwendungs- und SSH-Server nicht auf einem Rechner liegen. Auf diese Weise lässt sich ein Teilstück einer Strecke verschlüsseln, etwa der zwischen zwei Gateways. Damit auch andere Rechner den Tunnel benutzen können, müssen sie Zugriffsrecht auf *localhost:6080* besitzen. Das bewirkt die Option *–g* beim Einrichten des Tunnels.

In alle Richtungen

In die andere Richtung funktioniert das Remote Port Forwarding: Dort gibt man auf dem Gateway *gate*, der vor dem eigenen Intranet-Webserver sitzt, mit *ssh –R 6080:<intranetweb>:80 <remotegate>* den Intranet-Webserver als Ziel-Host an, aber den entfernten Gateway – etwa den einer Zweigstelle – als SSH-Server. Der lauscht dann auf Port 6080 auf Anfragen, die er an den SSH-Client auf *gate* weiterleitet, der sie wiederum an Port 80 des Intranet-Webservers weiterreicht. Damit nun alle Browser hinter dem Remote-Gateway auf den Intranet-Server zugreifen können, muss sich in der Datei */etc/ssh/sshd_config* auf *remotegate* die Zeile *GatewayPorts yes* befinden.

Hat man es mit mehreren hintereinanderliegenden Firewalls zu tun, kann man die Reichweite der SSH-Tunnel dadurch erhöhen, dass man mehrere ineinander verschachtelt – jeder weitere erhöht die Reichweite um zwei Hops. Dazu ist es notwendig, den Ziel-Port der äußeren Tunnel auf den zuvor geschaffenen Einstiegspunkt der inneren mit der Option *–p* umzulenken. Als Beispiel dienen zwei Tunnnel, an denen insgesamt sechs Rechner beteiligt sind. Zuerst etabliert man die Weiterleitung zwischen *host2* und *host5*, indem man auf *host3* den Befehl *ssh –gL 2222:host5:22 host4* absetzt. Danach benutzt man für das *ssh*-Kommando auf *host2* den neuen *ssh*-Tunnel-Port *2222* auf *host3* für den Zugriff auf den *sshd* auf *host5* mit *ssh –p 2222 –gL 7080:host6:80 host3*. Nun wird der Web-Client *host1* beim Zugriff auf *host2:6080* auf den Webserver auf *host6* weitergeleitet.

Für Verbindungen zu Remote-Hosts, die man des Öfteren heimsucht, lohnt sich das Einrichten von Session-Konfigurationen in der Datei *~/.ssh/config*. Dort trägt man für jede Session (Host) den gewünschte Remote-Host, den zu verwendenden Benutzernamen und weitere Argumente ein, wie SSH-Version, X11-Enabling oder einen anderen SSH-Port. Auch die Einstellung zur Kompression ist dort gut aufgehoben.

```
Host ixhost
  HostName ixhost.heise.de
  User susanne
  Protocol 2 # Verhindert Fallback auf Version 1
  ForwardX11 yes
  Port 2222 # Spricht remote-sshd auf Port 2222 an
  Compression yes
```

Für ein passwortloses Anmelden sattelt man auf die Authentifizierung per RSA- oder DSA-Schlüssel um. Die Schlüsselpaare erzeugt der Befehl *ssh-keygen –t [rsa|dsa]* mit oder ohne Passphrase und hinterlegt sie standardmäßig in *~/.ssh/id_[rsa|dsa]* und *~/.ssh/id_[rsa|dsa].pub*. *ssh-copy-id –i ~/.ssh/id_[rsa|dsa].pub <remote-host>* kopiert die öffentlichen Schlüssel auf die Remote-Hosts und hängt sie dort jeweils an die Datei *~/.ssh/authorized_keys* an. Sollte *ssh-copy-id* nicht vorhanden sein, bildet das Kommando *cat ~/.ssh/*.pub | ssh <remote-host> 'cat >> .ssh/authorized_keys'* eine annehmbare Alternative.

Hat man die Schlüssel mit Passphrase erzeugt, müsste man nun bei jedem Login die Passphrase angeben. Diese Arbeit kann man dem *ssh-agent* überlassen. Er speichert die bei der lokalen Anmeldung einmal eingegebene Passphrase in einem gesonderten Cache. Zur Anmeldung starten kann man ihn beispielsweise durch einen Eintrag wie *test "$SSH_AUTH_SOCK" || exec ssh-agent $SHELL -c "ssh-add; exec $SHELL –login"* etwa in der Datei *~/.profile*. (sun)

Der Traum vom Fliegen

Schwerelos

Kai König

Konnten die Menschen im Altertum nur davon träumen, sich wie Vögel und Insekten in der Luft zu tummeln, ist das Fliegen heute eine Selbstverständlichkeit und behördlich überwacht.

Eine kurze Recherche zur Geschichte der Luftfahrt fördert Interessantes zutage. Eine der frühesten Erwähnungen des Traums vom Fliegen findet sich in einer Legende über den chinesischen Kaisers Shun (2258-2208 v. Chr.), in der eben dieser Kaiser lernte, wie ein Vogel zu fliegen, um aus einer Gefangenschaft zu entkommen (www.luftrettung-hamburg.de/html/pioniere.html#Der%20chinesische%20Kaiser%20Shun) – berichtet auf den Webseiten der Luftrettung Hamburg. Neben vielen aktuellen Informationen zur Luftrettung und deren Geschichte bietet dieses Webangebot eine aus vier Teilen bestehende Übersicht über die Geschichte des Fliegens (www.luftrettung-hamburg.de/html/geschichte_des_fliegens.html).

Nachdem man im Mittelalter die Fähigkeit des Fliegens mit mystischen Begebenheiten assoziierte und Personen, denen diese Begabung nachgesagt wurden, oftmals auf dem Scheiterhaufen endeten, brachte die Renaissance die ersten wissenschaftlichen und methodischen Annäherungen an das Fliegen. Hier sind vor allem Leonardo da Vincis Entwürfe verschiedener Fluggeräte zu nennen (lrh10.fh-bielefeld.de/Projekte/Leonardo/start/modelle.htm), speziell der Fallschirm, die Luftschraube sowie der Gleiter.

Die ersten flugfähigen Modelle entstanden gegen Ende des achtzehnten Jahrhunderts und wurden im neunzehnten durch Fluggeräte abgelöst, die auf den Prinzipien der Aerodynamik basierten und Gleitflug ermöglichten. Pioniere dieser Epoche sind Otto Lilienthal und George Cayley. Den ersten belegbar motorisierten und gesteuerten Flug führten die Gebrüder Wright im Dezember 1903 durch und läuteten damit das Zeitalter der modernen Luftfahrt ein (de.wikipedia.org/wiki/Gebr%C3%BCder_Wright).

Warum heutzutage tonnenschwere Flugzeuge überhaupt abheben können, erklärt www.erklaert.de/warum/fliegen.htm anschaulich und erläutert die Konzepte von Auftrieb, Strömung, Steuerung und Navigation von Flugzeugen. Eine weitere empfehlenswerte Sammlung von Informationen zu diesem Thema findet sich im Wissensportal des SWRs (www.wissen.swr.de/warum/fliegen/themenseiten/t_index/s1.html).

Im Laufe der knapp mehr als hundert Jahre seit dem ersten motorisierten und gesteuerten Flug ist die Welt der Fliegerei natürlich nicht stehen geblieben. So hat sich im Bereich der zivilen Luftfahrt neben der kommerziellen Passagier-, Fracht- oder Postbeförderung eine lebendige und aktive Gemeinschaft von Hobbyfliegern etabliert. Da auch der Luftraum immer geschäftiger und verstopfter wird, hat man die zivile Luftfahrt vielen Regelungen unterworfen.

Luftige Standards

Eine der zentralen Organisationen im Bereich der Zivilluftfahrt ist die International Civil Aviation Organization (ICAO), eine Unterorganisation der Vereinten Nationen (www.icao.int). Eine ihrer Hauptaufgaben ist Standardisierung und Sicherheit im internationalen Flugverkehr. Dazu gehören unter anderem Fragen der Lizenzierung von Kabinenpersonal und Piloten im Rahmen der internationalen Flugabwicklung.

Daneben hat nahezu jedes Land lokale Luftfahrtbehörden, die für die praktische Umsetzung der zahlreichen internationalen Vorgaben sorgen. In Deutschland übernimmt diese Rolle das Luftfahrt-Bundesamt (www.lba.de/cln_010/DE/Home/homepage__node.html__nnn=true). Das LBA ist wiederum Teil der Joint Aviation Authorities (JAA, www.jaa.nl), einem Zusammenschluss von 34 nationalen Luftfahrtbehörden in Europa. Andere bekannte Luftfahrtbehörden sind die Federal Aviation Administration (FAA) in den USA sowie die CAA in Großbritannien und vielen anderen Commonwealth-Staaten wie Australien und Neuseeland.

Um den Bogen zurück zum Traum vom Fliegen zu spannen, soll es im Weiteren darum gehen, was ein am Fliegen Interessierter tun muss, um einen der begehrten „Führerscheine fürs Flugzeug" zu ergattern.

Portale wie flugschule24.de oder flugschule-online.de bieten einen regional sortierten Überblick über Flugschulen in Deutschland, Österreich und der Schweiz, dabei findet man sowohl kommerzielle als auch Trainingsanbieter, die an Aero Clubs angegliedert sind.

Wer Fliegen nur als Hobby ausüben möchte, sollte sich an der sogenannten Privatpiloten-Lizenz orientieren, auch mit PPL abgekürzt. Mit dem Erwerb dieser Lizenz darf man in der Regel private Flüge durchführen, dafür allerdings kein Geld verlangen. Je nach landesbeziehungsweise *AA-spezifischen Bestimmungen ist es aber möglich, die anfallenden Kosten (Treibstoff, Flugzeugmiete et cetera) auf Piloten und Mitflieger umzulegen – frei nach dem Prinzip der Kostenbeteiligung.

Die etwas andere GPL

In Deutschland gibt es neben der PPL A (www.ppl-a.com) für Flugzeuge die Lizenztypen PPL H und PPL D für Hubschrauber und Ballon. Segelflieger (GPL) und Piloten von Ultraleichtfliegern (SPL) haben ebenfalls ihre eigenen Lizenzen und Ausbildungsprogramme. Wer die Fliegerei kommerziell betreiben möchte, oder gar den Berufswunsch Flugzeugführer bei einer der großen Passagier-Airlines hat, muss deutlich höhere Hürden in Form der Commercial Pilot License (CPL) oder Airline Transport Pilot License (ATPL) meistern. Einen ausführlichen und für Einsteiger in die Materie geeigneten Überblick über die verschiedenen Lizenztypen sowie Trainingsinhalte und Kosten bietet die Webseite des Air Alliance

Flight Centers am Airport Siegerland (www.air-alliance.de/fc/index.php).

Hat man endlich die begehrte Lizenz, eine Maschine gechartert und diese in der Luft, gilt es, das Flugziel zu bestimmen – in der Realität sollte der Pilot diese Frage natürlich bereits im Rahmen der Flugplanung am Boden beantwortet haben. Dazu bedarf es Kenntnissen in Navigation, dem Lesen und Verstehen von Flugkarten sowie der Fähigkeit, per Funk mit Tower, Bodenkontrolle und Luftraumüberwachung zu kommunizieren (www.tf.uni-kiel.de/~fp/fliegerei/ausbildung/sprechfunk/sprechfunk1.html). Einen Eindruck in die komplexe Materie von Flugkarten bietet die Seite www.airports.de, die freie Übersichtskarten mit Runway-Layouts und Circuit-Mustern bereitstellt.

Ein beispielhafter, genauerer Blick auf den Flugplatz Ober-Mörlen (ein kleines Dörfchen in Mittelhessen und der Geburtsort des Autors) mit der internationalen Airport-Kennung EDFP (www.airports.de/airport/airport,_flughafen,_flugplatz/germany/hessen/ober-moeerlen_-_flugplatz/details.html) liefert detaillierte Informationen zu Dimension, Ausrichtung und Beschaffenheit der Landebahn sowie eine Übersichtskarte des Circuit-Musters (oftmals mit dem Begriff „Platzrunde“ beschrieben). Schaut man sich ein wenig mehr auf airport.de um, fallen die im Vergleich zu EDFP geschäftig wirkenden Karten internationaler Flughäfen wie Frankfurt oder Düsseldorf ins Auge. Alle Karten auf dieser Seite sind jedoch als grobe Orientierung zu verstehen und nicht für die aktuelle Anflugnavigation geeignet, hierfür sollten Piloten immer die aktuelle Version des offiziellen Kartenmaterials zur Verfügung haben.

Eine etwas andere Art, das Fliegen zu erlernen, bietet das Gymnasium Laucha in Sachsen-Anhalt (www.gymnasium-laucha.de/?id=104255000032). Diese Schule hat einen Schwerpunkt auf Luftsport sowie Luft- und Raumfahrt, und Schüler können dort das schuleigene Segelflugzeug nutzen, in verschiedenen AGs mitarbeiten sowie neben dem Unterricht eine Privatpilotenlizenz erwerben. Theorie und Praxis des Luftsports und des Fliegens sind in Wahlpflichtkurse in den Jahrgangsstufen 7 bis 12 in den Unterricht eingebunden. (ka)

URLs auf einen Blick

www.luftrettunghamburg.de/html/pioniere.html#Der%20chinesische%20Kaiser%20Shun

lrh10.fh-bielefeld.de/Projekte/Leonardo/start/modelle.htm

de.wikipedia.org/wiki/Gebr%C3%BCder_Wright

www.wissen.swr.de/warum/fliegen/themenseiten/t_index/s1.html

www.icao.int

www.lba.de/cln_010/DE/Home/homepage__node.html__nnn=true

www.jaa.nl

www.flugschule24.de

www.flugschule-online.de

www.ppl-a.com

www.air-alliance.de/fc/index.php

www.tf.uni-kiel.de/~fp/fliegerei/ausbildung/sprechfunk/sprechfunk1.html

www.airports.de

www.airports.de/airport/airport,_flughafen,_flugplatz/germany/hessen/ober-moeerlen_-_flugplatz/details.html

www.gymnasium-laucha.de/?id=104255000032

Wer weitere URLs zum Thema kennt, hat die Möglichkeit, sie der Online-Version (www.heise.de/ix/artikel/2008/01/152/) hinzuzufügen.

Vor 10 Jahren: Im kleinen Kreis

Videokonferenzen könnten Märkte zu Telegesprächen machen. Und ein Marktsegment bilden, in dem wirklich noch Geld verdient wird.

Zum Jahresanfang 1998 gründeten Ernst Malmsten und Kajsa Leander in London die Firma Boo.com, ein Internet-Modehaus. Die Geschäftsidee war etwas vage: Die Boo-Nutzerin sollte Klamotten auf einem Online-Püppchen (Avatar) ausprobieren und danach bei „Miss Boo“ das bestellen, was ihr am besten gefällt. Boo.com wurde mit der Gründung im Januar 1998 zum Superstar der Internet-Geschäftswelt ausgerufen – und war später die erste Firma, die in der „Dotcom-Blase“ spektakulär platzte. Von den Möglichkeiten des Internets berauscht, formulierte man Thesen wie „Märkte sind Gespräche“, die später sogar als Manifest die Runde machten.

Doch wer Modepüppchen anziehen, wer miteinander über das Netz ins Gespräch kommen will, braucht gute Verbindungen und jede Menge Software. Was mit Videosystemen wirklich möglich ist, stellte *iX* in Heft 1/1998 vor, das Software wie Hardware von Videokonferenzen unter die Lupe nahm. Unter dem arbeitswissenschaftlichen Begriff CSCW (Computer Supported Co-operative Work) wurden realistische Anwendungen vorgeführt. Dabei setzte man nicht einfach auf „das Internet“, sondern auf fortschrittliche Varianten wie MBone (IP Multicast Backbone) und das Breitband-Forschungsnetz B-WIN: Videokonferenzen, in denen ernsthaft gearbeitet wurde, waren etwas anderes als das aufkommende Instant Messaging mit Bildchen von der Webcam.

iX analysierte ein Beispiel aus der Praxis, bei dem Kapitän, Hafenmeister, Versicherungsagenten und Reeder eines havarierten Schiffes über eine ATM-Verbindung zwischen Antwerpen und Bremen darüber debattierten, ob das Schiff noch den nächsten Hafen anlaufen kann. Die aufwendige Videokonferenz, komplett mit der Einspielung von Videoaufnahmen vom Unfall und der gemeinsamen Arbeit an einer Riss-Zeichnung des Schiffes in einem Whiteboard-Tool, wurde wissenschaftlich wie ökonomisch ausgewertet. 18 000 DM kostete allein die ATM-Leitung der dreistündigen Konferenz, denen 20 000 DM Liegegebühren für das beschädigte Schiff und 30 000 DM Reisekosten für die diversen Sachverständigen gegenüberstanden. Fazit: Videokonferenzen rechnen sich, schonen Geldbeutel und Umwelt. Sie sind aber nur sinnvoll, wenn die Technik einwandfrei funktioniert und alle Beteiligten „Entscheider“ sind, die nicht erst ein Okay von Vorgesetzten einholen müssen.

Auch heute sind Videokonferenzen nicht billig. „Telepresence“, ein System, das T-Systems in Kooperation mit Cisco vermarktet (s. S. 113), kommt für sechs Personen auf 300 000 €, zuzüglich Kosten für eine Standleitung mit mindestens 15,5 MBit/s. Wer an den halbrunden Tischen von „Telepresence“ sitzt, fühlt sich an die „privaten Fernsehstudios“ erinnert, die Firmen wie Daimler oder Siemens bis in die 90er-Jahre unterhielten. Telepresence wird aber nicht von Konzernen eingesetzt, sondern vor allem von mittelständischen Unternehmen. Sie wollen nicht länger Mitarbeiter auf zunehmend umständlicher werdende Flugreisen schicken. Märkte sind Gespräche, aber nur, wenn man nicht von der Technik behindert wird. *Detlef Borchers*

MEHR KBYTES

Datenbanken

Nicht zuletzt dank der Tatsache, dass viele preiswerte Angebote von Internet-Service-Providern heutzutage die Benutzung einer Datenbank beinhalten, können und müssen sich die Webverantwortlichen in den Umgang mit Tabellen und SQL einarbeiten, um den Besuchern der eigenen Website die richtigen Daten anbieten zu können.

Dass solche dynamischen Webseiten, die bei jedem Aufruf aktuelle Daten beim Server erfragen, außerdem eine Skriptsprache erfordern, in der die Datenbankanfragen formuliert sind, kommt erschwerend hinzu (soll aber hier unberücksichtigt bleiben).



SQL, die Structured Query Language, erweist sich schnell als eine vergleichsweise leicht erlernbare Sprache, denn der Sprachschatz hält sich in engen Grenzen. Allerdings unterscheiden sich die Datenbanksysteme hinsichtlich ihrer SQL-Implementierung, sodass nicht gewährleistet ist, dass in MySQL korrektes SQL in Oracle und anderen Systemen ebenfalls richtig ist.

Marcus Throll und Oliver Bartosch haben aus diesem Grund ihrem bei Galileo veröffentlichten „Einstieg in SQL“ eine Gegenüberstellung der Syntax für eine Reihe von Datenbanksystemen beigegeben. Das Buch richten die Autoren an Einsteiger wie Erfahrene, Letztere können gelegentlich auf einführende Passagen verzichten. Am Anfang steht der DB-Entwurf (bis zur fünften Normalform und zum Entity Relationship Model) in gebotener Kürze. Es folgt ein Kapitel zur Tabellendefinition, bevor es an *insert*, *select* et cetera geht. Für Erfahrene behandeln Throll und Bartosch weiterreichende Themen wie Transaktionen oder *grant* und *revoke*. Der CD-ROM haben die Autoren eine selbstentwickelte Lernsoftware beigelegt (nur für Windows).

Lynn Beighleys „Head First SQL“ kommt ohne theoretisches Vorabkapitel aus. Das liegt sicherlich daran, dass dieser Band Bestandteil der gleichnamigen Buchreihe (ohne „SQL“) ist, die durchweg mit einer Mischung aus Comic, Fotos, gleichsam Handschriftlichem, Text et cetera versucht, das Hirn des Lesenden aus der Reserve zu locken. Wer zum ersten Mal eins der Bücher aufschlägt, dürfte das Layout und die Aufbereitung des Inhalts gewöhnungsbedürftig finden, aber eine abwechslungsreiche Einführung kommt dabei meist heraus. Von der Schwierigkeit her reicht der Band bis zu den unterschiedlichen *joins* und Transaktionen. Die deutsche Übersetzung „SQL von Kopf bis Fuß“ erscheint Anfang 2008. Im Rahmen einer Sonderaktion „3 für 2“ (www.oreilly.de/3fuer2) können Vielleser dies Werk bis zum Frühjahr billiger bekommen, wenn sie genug Interesse an so vielen Bänden haben.

Im selben Verlag erschienen ist Andrew Cummings und Gordon Russells „SQL Hacks“ (hinter „Hacks“ steckt wiederum eine Buchreihe), das die Verlagswebsite für 2006 ankündigt, während der Band selbst das Ende 2007 vorgibt (richtig, heißt es, ist das ältere Datum). Anders als die beiden vorgenannten Einführungen dürfte diese Einhundert-Tipp-Sammlung Bonbons für erfahrene SQLer enthalten. Und um gleich bei O’Reilly zu bleiben: Des RDBMS-Gurus C. J. Date knappe Darstellung relationaler Datenbanken liegt jetzt in deutscher Sprache vor – in Vokabelheftgröße, mithin definitiv urlaubsgeeignet.

Zu einzelnen Datenbanksystemen haben diverse Verlage Bücher im Programm, manche gleich Serien. So hat Addison-Wesley 2007 Buck Woodys „Administrator’s Guide to SQL Server 2005“ unter dem Titel „SQL Server 2005“ veröffentlicht – mit einer 180-Tage-Testversion der deutschen Enterprise-Edition. Zwar steht die 2008er Version schon fast vor der Tür, aber konservative DBAs dürften noch eine Weile mit der jetzt vorhandenen Release arbeiten wollen (bis Service Pack X). Hier geht es naturgemäß hauptsächlich um die Architektur des konkreten DBMS-Servers, wie er zu installieren und zu warten ist, wie man ihn optimiert et cetera. Nichts für Einsteiger.

Als kompakten Einstieg hat die Bonner Redline-Tochter mitp Stefan Heitsieks ebenfalls konkretes „Oracle Express Edition“ bezeichnet. Gegen die gut 650 Seiten SQL Server kommen Heitsieks 440 geradezu schmal daher. Sie eignen sich für all diejenigen, die die kostenlose Oracle-Version ausprobieren und sich weiterbilden wollen. Logischerweise handelt es sich nicht nur um eine SQL-Einführung (circa 20 Seiten).

In seiner Buchreihe von OpenOffice-Bänden hat Galileo Computing jetzt die zweite Auflage von Thomas Krumbeins „Datenbanken mit OpenOffice.org 2.3“ zum Modul Base und dem Datenbanksystem HSQLDB der freien Software veröffentlicht. Die Einführung, zu der der gesondert zu installierende Report Designer gehört, sollte Datenbankneulingen den Einstieg erleichtern – unterstützt durch eine Videoschulung auf der DVD.

Henning Behme

Lynn Beighley; Head First SQL; Sebastopol, CA (O’Reilly Media) 2007; 571 Seiten; € 43,00 (Paperback) – Anfang 2008 erscheint die deutsche Ausgabe: **SQL von Kopf bis Fuß**; übersetzt von Lars Schulten; € 49,95 (Paperback)

Andrew Cumming, Gordon Russel; SQL Hacks; Tips & Tools for Digging into Your Data; Sebastopol, CA (O’Reilly Media) 2007; 386 Seiten; € 29,00 (Paperback)

C. J. Date; Fachwörterbuch Relational Datenbank; Köln (O’Reilly) 2007; 118 Seiten; übersetzt von Dorothea Heymann-Reder; € 9,90 (Paperback)

Stefan Heitsiek; Oracle express Edition; Heidelberg (mitp/Redline) 2007; 2., aktualisierte und erweiterte Auflage; 412 Seiten zzgl. CD-ROM; € 34,95 (Paperback)

Thomas Krumbein; Datenbanken mit OpenOffice.org 2.3; Bonn (Galileo Computing) 2007; 547 Seiten zzgl. DVD; € 39,90 (gebunden)

Marcus Throll, Oliver Bartosch; Einstieg in SQL; Verstehen, einsetzen, nachschlagen; Bonn (Galileo Computing) 2007; 2., und aktualisierte erweiterte Auflage; 291 Seiten zzgl. CD-ROM; € 24,90 (gebunden)

Buck Woody; SQL Server 2005; Das Handbuch für Administratoren; München (Addison-Wesley) 2007; 662 Seiten zzgl. DVD; übersetzt von G & U; € 49,95 (gebunden)



Jürgen Plate

Linux Hardware Hackz

Messen, Steuern und Sensorik mit Linux

München, Wien 2007
Carl Hanser
462 Seiten
39,90 €
ISBN 978-3-446-40783-1

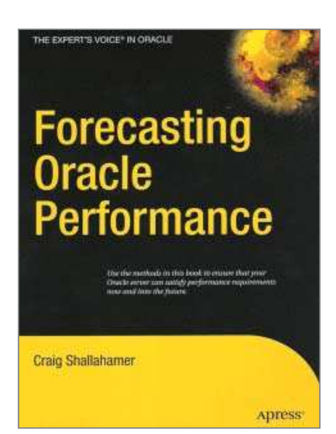
Zu den wichtigen Eigenschaften von Linux zählen Flexibilität, Stabilität, Geschwindigkeit, sparsamer Umgang mit Systemressourcen, offene Quelltexte und eine große Community. Gründe genug dafür, dass Linux inzwischen nicht nur als Desktop- und Serverbetriebssystem verwendet, sondern immer häufiger für eingebettete Systeme eingesetzt wird. Dabei gelten solche Systeme seit jeher als anspruchsvoll. Doch nicht aller Anfang muss schwer sein, wie das vorliegende Buch zeigt.

Profitieren dürften von diesem Band vor allem Entwickler eingebetteter Systemen sowie Studenten und Technikbegeisterte aus den Fachbereichen Elektrotechnik und Informatik. Es enthält alles Wissenswerte für den Einstieg in die Welt des Messens, Steuerns und Regelns, wie folgender Themenabriss zeigt: Nach einer knappen Einführung stellt Plate exemplarisch eine eigene abgespeckte Linux-Distribution zusammen, die als Basis für weitere Experimente dienen kann. Eine Einführung in die Systemprogrammierung unter Linux folgt. In den nächsten Kapiteln geht der Autor Stück für Stück auf die Arbeit mit den Standard-PC-Schnittstellen (Parallelport, RS232, Gameport, USB) ein und erläutert die verschiedenen Chip-Schnittstellen SPI, I^2C und 1-Wire. Danach beschreibt er die digitale Ein- und Ausgabe mit Tastern, Schaltern, Optokopplern, Relais, Verstärkerschaltungen et cetera. Es folgen Kapitel über Sensoren, Motorsteuerung, Anzeigen und Display und Fernsteuerung über Infrarot, Funk und Handy. Zudem finden sich umfangreiche Exkurse zur Analog-Digital-Wandlung, zur Messung elektrischer Größen und zur Datenauswertung.

Sprache, Layout und (einfarbige) Bebilderung tragen zum Lesevergnügen bei. Neben Codebeispielen finden sich zahlreiche Schaltungsentwürfe im Buch, die von der Website als Eagle-Dateien heruntergeladen werden können. Mit diesen Dateien kann man die vorhandenen Schaltungen nach eigenen Wünschen verändern. Zusätzlich enthält die Website weiterführende Dokumente und ein Extra-Kapitel zum Download.

Alles in allem bereitet das Buch viel Spaß und lädt ein, sich mit der Materie auseinanderzusetzen. Allerdings sollten Leser gehörige Elektronik-Grundkenntnisse mitbringen und schon den einen oder anderen Quelltext zu Gesicht bekommen haben.

FLORIAN POTSCHKA



Craig Shallahamer

Forecasting Oracle Performance

Berkeley, CA 2007
Apress
269 Seiten
39,99 US-$
ISBN 978-1-59059-802-3

Einer speziellen, in der Praxis jedoch eher vernachlässigten Frage widmet sich der Oracle-Experte Shallahamer: Kann man das Wechselspiel zwischen CPU-und I/O-Verhalten eines Datenbanksystems modellhaft beschreiben und Vorhersagen treffen? Und wie lässt sich der Workload quantitativ ermitteln? Wie Shallahamer in der Einleitung schreibt, ist dies ein Gebiet, das selbst gestandene Kapazitätsplaner in Projekten zur Kapitulation zwingen kann. Das vorliegende dürfte zu den wenigen Oracle-Büchern gehören, die nicht an eine bestimmte Release gebunden sind. Im Grunde ist der Inhalt nicht datenbankgebunden, da die Kernthemen CPU, I/O und Workload herstellerunabhängig sind. Insgesamt ist „Forecasting Oracle Performance“ unterhaltsam geschrieben, flüssig zu lesen und sämtliche Beispiele orientieren sich an bekannten Fragestellungen aus der Praxis. Es „will“ kein Lehrbuch sein, sondern eine Art Trainingskurs, wie der Autor einleitend anmerkt.

Im ersten Kapitel erläutert er, warum es wichtig ist, ein passendes Modell zur Beschreibung zu erstellen. Anschließend beleuchtet er die elementaren Vorgänge bei Transaktionen anhand von Themen wie „Arrival Rate“, „Service Time“ und „Reponse Time“ näher. Leider kommt das Thema Datenermittlung zu kurz. Gerade bei der für den Leser spannenden Frage, welche Daten im Oracle-Umfeld gesammelt werden sollten, um den Workload zu charakterisieren, hält sich der Autor bedeckt. Zwar greift er dieses Thema in Kapitel 7 nochmals auf und erläutert im Text ein Shell-Script zur CPU-Auslastung auf Unix-Ebene mittels SAR, doch welche Daten konkret bei einem Oracle-Server gesammelt werden sollten, bleibt offen. Shallahamer belässt es bei allgemeinen Hinweisen, dass es eben bei einem Oracle-Server viele Parameter und Variablen (in V$-Views) gibt. Insgesamt sind die mathematischen Ausführungen bewusst knapp gehalten, wie der Autor immer wieder anmerkt. Da es sich meist um einfache Formeln handelt, verlangt er dem Leser kein tieferes Verständnis ab. So gerüstet mit einfachen Modellannahmen, lassen sich wichtige Fragestellungen erstaunlich gut beantworten: Was passiert, wenn man mehr CPU-Kapazität hinzufügt oder CPUs einsetzt, die beispielsweise 20 % schneller sind. Im mathematisch anspruchsvolleren Kapitel 3 erläutert der Autor den Grundgedanken der „Erlang C“-Warteschlangen-Theorie.

Wie man Daten statistisch mittels Regressionsanalysen auswertet, zeigt Kapitel 4. Einer Variante, der linearen Regression, ist ein weiteres Kapitel gewidmet. Viele Fallbeispiele anhand der Warteschlangentheorie und Variation von Parametern zeigt das fünfte Kapitel. Das siebte zum Thema „Characterizing the Workload“ schließt mit dem Hinweis, dass eine modellhafte Beschreibung großer Oracle-Systeme schnell an ihre Grenzen stößt. Das Buch endet mit einem Kapitel zum Thema „Scalability“ (Amdahl’s law).

DR. WOLFGANG GABRIEL



Andy Budd, Cameron Moll, Simon Collins

CSS Mastery

Webdesign für Fortgeschrittene

München November 2006
Addison-Wesley
270 Seiten
39,95 €
ISBN 978-3-8273-2457-3

Titel zum Thema CSS sind Legion, dieser in der von Addison-Wesley als „Sign of Exellence“ beworbenen Reihe erschienene gehört zu den besseren (in derselben Serie ist Zeldmans „Webdesign mit Webstandards“ erschienen, siehe *iX* 4/07). Es richtet sich an den fortgeschrittenen Webdesigner, daher setzt es ein Grundwissen über CSS und (X)HTML voraus.

Im zweiten Kapitel geht Andy Budd auf das Boxmodell ein; selbstredend dass er die Eigenheiten des IE 5.x und IE 6 für Windows thematisiert, da diese Browser im Quirks-Modus sich leider nicht standardkonform verhalten. Es fällt in diesem Zusammenhang auf, dass die Screenshots fast ausnahmslos vom Safari unter Mac OS X stammen, denn der Autor verhehlt seine Affinität zum Mac nicht. Ausführlich handelt er unter anderem das Thema Positionierung mit CSS ab.

In den weiteren Kapiteln erfährt der Leser, wie er Boxen mit abgerundeten Ecken und Schlagschatten für Bilder erstellt, Links ansprechend formatieren kann oder ansprechende Navigationsleisten erzeugt.

Ein Kapitel widmet Budd dem Formatieren von Formularen und Datentabellen. Dabei zeigt er, wie man es nicht machen sollte, wenn Tabellen leserlich und übersichtlich sein sollen. Barrierefreiheit zieht sich wie ein roter Faden durch das Werk.

Layouttechniken wie Zentrieren, negative Ränder oder schwimmendes Layout kommen ebenfalls zu ihrem Recht, ehe es um das wirkliche Webdesigner-Leben geht: Hacks, Filter, Bugs und Fehlerbehebung. Die vorgestellten Workarounds beziehen sich aber allesamt auf die Windows-Browser IE 5.x und IE 6.

Wer Anmerkungen zum IE 7 erwartet, wird leider enttäuscht, denn das amerikanische Original ist zu alt, als dass es ihn behandeln könnte, lediglich einige Praxisbeispielen erwähnen ihn kurz. Die Übersetzung ist gelungen und lässt sich flüssig lesen, allerdings sieht man den Codebeispielen von Budd an, dass er sie nie ausprobiert hat. Es haben sich einige Tippfehler eingeschlichen, was der erfahrene Webdesigner aber geflissentlich überlesen kann. Die Koautoren Simon Collins und Cameron Moll werten mit ihren Praxisbeispielen das Werk auf und zeigen anschaulich, wie man das in der Theorie vorgestellte Wissen in Webprojekten umsetzen kann.

KARSTEN KISSER



MAGAZIN FÜR PROFESSIONELLE INFORMATIONSTECHNIK

Postfach 61 04 07, 30604 Hannover; Helstorfer Straße 7, 30625 Hannover

Redaktion
Telefon: 05 11/53 52-387, Fax: 05 11/53 52-361, E-Mail: post@ix.de
Abonnements: Telefon: 051 37/88 20 00, Fax: 051 37/88 17 12, E-Mail: abo@heise.de

Herausgeber: Christian Heise, Ansgar Heise

Redaktion: Chefredakteur: Jürgen Seeger (JS) -386

Stellv. Chefredakteur: Henning Behme (hb) -374

Ltd. Redakt.: Kersten Auel (ka) -367, Ralph Hülsenbusch (rh) -373, Bert Ungerer (un) -368

Jürgen Diercks (jd) -379, Christian Kirsch (ck) -590, Wolfgang Möhle (WM) -384, Susanne Nolte (sun) -689, André von Raison (avr) -377, Michael Riepe (mr) -787, Ute Roos (ur) -535

Redaktionsassistenz: Carmen Lehmann (cle) -387, Michael Mentzel (mm) -153

Korrespondent Köln/Düsseldorf/Ruhrgebiet:
Achim Born, Siebengebirgsallee 82, 50939 Köln,Telefon: 02 21/4 20 02 62,
E-Mail: ab@ix.de

Korrespondentin München:
Susanne Franke, Ansbacherstr. 2, 80796 München, Telefon: 089/28 80 74 80,
E-Mail: sf@ix.de

Ständige Mitarbeiter: Torsten Beyer, Detlef Borchers, Fred Hantelmann, Kai König, Michael Kuschke, Barbara Lange, Stefan Mintert, Holger Schwichtenberg, Susanne Schwonbeck, Christian Segor, Diane Sieger, Axel Wilzopolski, Nikolai Zotow

DTP-Produktion: Enrico Eisert, Wiebke Preuß, Matthias Timm, Hinstorff Verlag, Rostock

Korrektorat/Chefin vom Dienst: Anja Fischer

Fotografie: Martin Klauss Fotografie, Despetal/Barfelde

Titelidee: ***iX;*** Titel- und Aufmachergestaltung: Dietmar Jokisch

Verlag und Anzeigenverwaltung:
Heise Zeitschriften Verlag GmbH & Co. KG, Postfach 61 04 07, 30604 Hannover;
Helstorfer Straße 7, 30625 Hannover; Telefon: 05 11/53 52-0, Fax: 05 11/53 52-129

Geschäftsführer: Ansgar Heise, Steven P. Steinkraus, Dr. Alfons Schräder

Mitglied der Geschäftsleitung: Beate Gerold

Verlagsleiter: Dr. Alfons Schräder

Anzeigenleitung: Michael Hanke -167, E-Mail: michael.hanke@heise.de

Assistenz: Christine Richter -534, E-Mail: christine.richter@heise.de

Anzeigendisposition: Christine Richter -534, E-Mail: christine.richter@heise.de

Anzeigenverkauf: PLZ-Gebiete 0–3, Ausland:
Oliver Kühn -395, E-Mail: oliver.kuehn@heise.de,
PLZ-Gebiete 8–9: Ralf Räuber -218, E-Mail: ralf.raeuber@heise.de
Sonderprojekte: Isabelle Paeseler -205, E-Mail: isabelle.paeseler@heise.de

Anzeigen-Inlandsvertretung: PLZ-Gebiete 4–7:
Karl-Heinz Kremer GmbH, Sonnenstraße 2, D-66957 Hilst,
Telefon: 063 35/92 17-0, Fax: 063 35/92 17-22, E-Mail: karlheinz.kremer@heise.de

Anzeigen-Auslandsvertretung:
Großbritannien, Irland: Oliver Smith & Partners Ltd. Colin Smith, 18 Abbeville Mews, 88 Clapham Park Road, London SW4 7BX, UK, Telefon: (00 44) 20/79 78-14 40, Fax: (00 44) 20/79 78-15 50,
E-Mail: colin@osp-uk.com

Anzeigenpreise: Es gilt die Anzeigenpreisliste Nr. 19 vom 1. Januar 2007.

Vertriebsleitung: Mark A. Cano (-299); Abo-Service, Telefon: +49 (0) 711/7252-292

Teamleitung Werbung: Julia Conrades (-156)

Teamleitung Herstellung: Bianca Nagel (-456)

CD-Herstellung: Klaus Ditze (Ltg.), Nicole Tiemann

Druck: Dierichs Druck + Media GmbH & Co. KG, Kassel

Sonderdruck-Service: Ruth Utesch (-359, Fax: -360)

Verantwortlich: Textteil: Jürgen Seeger; Anzeigenteil: Michael Hanke

***iX* erscheint monatlich**

Einzelpreis € 5,50, Österreich € 6,20, Schweiz CHF 10,70, Benelux € 6,70, Italien € 6,70

Das Abonnement für 12 Ausgaben kostet: Inland € 56,–, Ausland (außer Schweiz) € 61,–; Studentenabonnement: Inland € 42,00, Ausland (außer Schweiz) € 47,00 nur gegen Vorlage der Studienbescheinigung (inkl. Versandkosten Inland € 8,30, Ausland € 13,30), Luftpost auf Anfrage.

iX-Abo+ (inkl. jährlicher Archiv-CD-ROM) jeweils zzgl. € 8,–

Für GI-, VDI-KfIT-, GUUG-, IUG-, LUG-, AUGE- und Mac-e.V.-Mitglieder gilt der Preis des Studentenabonnements (gegen Mitgliedsausweis).

Kundenkonto in Österreich:
Dresdner Bank AG, BLZ 19675, Kto.-Nr. 2001-226-00 EUR, SWIFT: DRES AT WX

Kundenkonto in der Schweiz: UBS AG, Zürich, Kto.-Nr. 206 P0-465.060.0

Abo-Service:
Heise Zeitschriften Verlag, Kundenservice, Postfach 810520, 70522 Stuttgart,
Telefon: +49 (0) 711/7252-292, Fax: +49 (0) 711/7252-392, E-Mail: abo@heise.de

Für Abonnenten in der Schweiz Bestellung über:
Thali AG, Aboservice, Industriestr. 14, CH-6285 Hitzkirch,
Telefon: 041/919 66 11, Fax: 041/919 66 77, E-Mail: abo@thali.ch, Internet: www.thali.ch (Jahresabonnement: CHF 111,–; Studentenabonnement: CHF 83,25)

Das Abonnement ohne Archiv-CD-ROM ist jederzeit mit Wirkung zur jeweils übernächsten Ausgabe kündbar. Das *iX-Abo+* (incl. jährlicher Archiv-CD-ROM) gilt zunächst für ein Jahr und ist danach zur jeweils übernächsten Ausgabe kündbar.

Vertrieb Einzelverkauf (auch für Österreich, Luxemburg und Schweiz): MZV Moderner Zeitschriften Vertrieb GmbH & Co. KG, Breslauer Str. 5, 85386 Eching, Telefon: 089/31906-0, Fax: 089/31906-113, E-Mail: mzv@mzv.de, Internet:: www.mzv.de

ISSN 0935-9680





Nach Klang sortierte Daten

Self-organizing Maps sind zwar nicht neu, kommen aber gerade wieder in Mode. Die zu Beginn der Achtziger von Teuvo Kohonen „erfundene“ Art der Datenvisualisierung erlaubt es, mehrdimensionale Daten zu gruppieren und grafisch darzustellen, beispielsweise die Mozart-Karte, eine über den Kopf des Komponisten geblendete Insellandschaft.

Linux im Unternehmenseinsatz

Mit der Advanced Version seiner auf Debian basierenden Server-Distribution schickt sich Xandros an, im lukrativen Markt für Enterprise-Linux Fuß zu fassen. Dabei müssen sich die Kanadier gegen die Platzhirsche in diesem Segment – Red Hat und Novell – behaupten. Ein Vergleich mit dem gerade freigegebenen RHEL 5.1 sowie dem SLES 10 SP1 beleuchtet, wo Xandros die Nase vorn hat und wo die Entwickler noch nacharbeiten müssen.

Java-Entwicklung per RAD

Obwohl Java-Programmierung nicht gerade als einfach gilt und IDEs wie Netbeans und Eclipse vor allem den Neuling mit zahllosen Einstelloptionen überfordern, spielen Werkzeuge für das sogenannte Rapid Application Development in diesem Sektor bislang keine große Rolle. Die Firma Xdev will das mit ihrem gleichnamigen Werkzeug ändern. Sie verspricht vollmundig „Java-Anwendungen in Rekordzeit“.

Heft 2/2008 erscheint am 17. Januar 2008

Webprogramme für das iPhone

Bislang lässt Apple keine nativen Anwendungen außer den von ihm mitgelieferten auf dem iPhone zu. Nur im Browser laufende, mit Ajax dynamisierte Applikationen können den Funktionsumfang des Geräts erweitern. Hält man sich an ein paar Regeln, lassen sich leicht passende Programme für das Handy erstellen und mit ein paar Tricks sogar debuggen.



Display-Technik: Geräte und Verfahren

Projektoren und Großdisplays gehören inzwischen zur Standardausstattung für Präsentationen im Unternehmen. In der Regel kommen dafür Groß-Displays und Projektoren zum Einsatz. Eine Marktübersicht zeigt die zurzeit erhältlichen Produkte und beleuchtet die Vor- beziehungsweise Nachteile der angewandten Techniken.

Das bringen





720p-Projektoren für Heimkino, Spiel und Präsentation

c’t-Notfall-CD gegen Viren und andere PC-Probleme

Linux-Server für den Business-Bereich

UPnP erleichtert die Heimnetzkonfiguration

Heft 26/07 jetzt im Handel





Die „Physik“ der Börse: Wie Spitzen-Wissenschaftler die Märkte steuern

Akkus auf Rädern: Elektroautos der nächsten Generation vor dem Start

Pharma vs. Evolution: Der lange Weg zu Mitteln gegen Fettleibigkeit

Heft 12/07 jetzt im Handel

TELEPOLIS

MAGAZIN DER NETZKULTUR



Raffael Schuppisser: Computerspiele sind Erzählungen

Bernd Schröder: Schleichender Störfall Antihaftbeschichtungen

www.heise.de/tp/